# MATHEMATISCHE ANNALEN

BEGRÜNDET 1868 DURCH

ALFRED CLEBSCH UND CARL NEUMANN

FORTGEFÜHRT DURCH

FELIX KLEIN — DAVID HILBERT

OTTO BLUMENTHAL — ERICH HECKE

GEGENWÄRTIG HERAUSGEGEBEN VON



122. Band · 1. Heft

(ABGESCHLOSSEN AM 10. MAI 1950)

BERLIN · GÖTTINGEN · HEIDELBERG
SPRINGER-VERLAG
1950

Math. Ann.

Math. Annalen, Bd. 122, S. 1—5 (1950).

# Über einen Satz der Theorie der Baireschen Funktionen und Borelschen Mengen.

Von

**Demetrios A. Kappos** in Erlangen.

In der Theorie der Baireschen Funktionen auf einer Menge $\mathfrak{M}$ und der dazu in engster Beziehung stehenden Borelschen Mengen in $\mathfrak{M}$ spielt folgender Satz eine Hauptrolle:

Ist $\Phi$ (bzw. $R$) ein Verband von Funktionen auf $\mathfrak{M}$ (bzw. von Mengen in $\mathfrak{M}$), so gilt:

$$\Phi_\lambda = \Phi_{gh} \cup \Phi_{hg} \text{ (bzw. } R_\lambda = R_{\sigma\delta} \cap R_{\delta\sigma}) \text{ [1]}$$

Der Satz wird bekanntlich[2]) zuerst für Funktionenverbände bewiesen und dann mit Hilfe der charakteristischen Funktionen der Mengen auf Mengenverbände übertragen. Sierpinski[3]) hat den Satz für Mengenverbände auch direkt bewiesen. In beiden Fällen aber wird beim Beweise der Punktbegriff benutzt, also wesentlich die Tatsache daß $R$ ein Mengenverband ist. Der Satz ist jedoch rein verbandstheoretischer Natur. In der Tat sind alle Begriffe, mit Hilfe derer der Satz formuliert wird, verbandstheoretisch erklärbar und dementsprechend gilt, wie wir zeigen werden, der Satz in topologischen Verbänden, die viel allgemeiner als die Funktionen- (bzw. Mengen-) Verbände sind.

1. *Algebraische Konvergenz in Verbänden.* In jedem Verein (teilweise geordnete Menge) $V$ kann man auf Grund der Vereinsrelation $\subseteq$ eine Konvergenz, sog. ($o$)-Konvergenz, einführen[4]), nämlich:

(1.1.) *Definition. Eine Folge $x_1, x_2, \ldots$ aus $V$ konvergiert gegen $x \in V$, wenn zwei Folgen $(a_n)$ und $(b_n)$ aus $V$ derart existieren, daß* I) $a_n \subseteq a_{n+1} \subseteq \subseteq x_{n+1} \subseteq b_{n+1} \subseteq b_n$ *für alle* $n = 1, 2, \ldots$ *und* II) $\cup\, a_n = \cap\, b_n = x$ *gilt.*

Es sei $V$ ein bedingter ($\sigma$, $\delta$)-Verband, d. h. in $V$ sei die $\sigma$ bzw. $\delta$ Operation für die in $V$ beschränkten abzählbaren Folgen von Elementen immer ausführbar; dann ist in $V$ die Def. (1.1) gleichwertig mit:

(1.2). *Eine beschränkte Folge $x_1, x_2, \ldots$ aus $V$ konvergiert gegen $x \in V$, wenn* $x = \bigcup\limits_{\nu=1}^{\infty} \bigcap\limits_{\varrho=0}^{\infty} x_{\nu+\varrho} = \bigcap\limits_{\nu=1}^{\infty} \bigcup\limits_{\varrho=0}^{\infty} x_{\nu+\varrho}$ *gilt.*

---

[1]) Sierpinski, W.: Fund. Math. **18**, 1--22 (1932). Den von Sierpinski benutzten Begriff „*Ring*" haben wir mit dem gleichbedeutenden Begriff „*Verband*" ersetzt, um besonders bei Funktionen Verwechslungen mit dem Begriff des algebraischen Ringes zu vermeiden. $\Phi_g$ bzw. $\Phi_h$ bzw. $\Phi_\lambda$ entsteht aus $\Phi$ durch Adjunktion der Limiten aller auf- bzw. absteigenden bzw. beliebigen eigentlich konvergenten Folgen aus $\Phi$. $\Phi_{gh}$ bzw. $\Phi_{hg}$ ist dann $= (\Phi_g)_h$ bzw. $= (\Phi_h)_g$.

[2]) Hahn, H.: Reelle Funktionen, S. 248 und 250, Leipzig 1932; auch Sierpinski l. c. 1)

[3]) Sierpinski, W.: C. r. Paris **192**, 1626—1628 (1931); auch Hahn l. c. 2).

[4]) Birkhoff, G.: Lattice Theory, § 36, New York 1940.

Wir werden im folgenden die ($o$)-Konvergenz *algebraisch*, in Zeichen $\lim \operatorname{alg} x_n = x$, nennen[5]). In einem bedingten ($\sigma$, $\delta$)-Verband $V$ existieren für eine in $V$ beschränkte Folge von Elementen $a_1, a_2, \ldots$

$$\underline{\lim} \operatorname{alg} a_n = \bigcup_{\nu=1}^{\infty} \bigcap_{\varrho=0}^{\infty} a_{\nu+\varrho} \text{ bzw. } \overline{\lim} \operatorname{alg} a_n = \bigcap_{\nu=1}^{\infty} \bigcup_{\varrho=0}^{\infty} a_{\nu+\varrho},$$

und es ist allgemein $\underline{\lim} \operatorname{alg} a_n \subseteq \overline{\lim} \operatorname{alg} a_n$ und im Falle der Konvergenz wegen (1.2) $\underline{\lim} \operatorname{alg} a_n = \overline{\lim} \operatorname{alg} a_n = \lim \operatorname{alg} a_n$.

2. *Topologische Verbände.* Es sei $V$ ein bedingter ($\sigma$, $\delta$)-Verband und $W$ irgendein Unterverband von $V$. Wir nennen $W$ *topologisch* bezüglich der algebraischen Konvergenz[6]) (kurz topologisch), wenn folgendes gilt:

(2.1). *Aus* $\lim \operatorname{alg} a_n = a$ *und* $\lim \operatorname{alg} b_n = b$; $a_n, b_n, a, b \in W$, $n = 1, 2, \ldots$ *folgt* $\lim \operatorname{alg} (a_n \cup b_n) = a \cup b$ *und* $\lim \operatorname{alg} (a_n \cap b_n) = a \cap b$.

Aus dieser Definition folgt:

(2.2). *Jeder Unterverband eines topologischen Verbandes ist wieder topologisch.*

*Anmerkung.* Die gewöhnliche Limesbildung von Mengenfolgen[7]) bzw. von Funktionenfolgen, die auf einer festen Menge $\mathfrak{M}$ definiert sind, ist, verbandstheoretisch aufgefaßt, ein spezieller Fall der algebraischen Limesbildung und Mengen- (bzw. Funktionen-)verbände sind bezüglich dieser Limesbildung topologisch. Die Mengen-(bzw. Funktionen-)Verbände sind jedoch distributiv, was für topologische Verbände allgemein nicht der Fall ist.

3. Es sei $V$ ein topologischer bedingter ($\sigma$, $\delta$)-Verband und $W$ irgendein Unterverband von $V$. Wir bezeichnen mit $W_\lambda$ das Untersystem von $V$, das wir erhalten, wenn wir zu $W$ alle Elemente aus $V$ adjungieren, die algebraische Limiten von Folgen aus Elementen von $W$ sind. Aus der Erklärung der Elemente von $W_\lambda$ und aus (2. 1) folgt, daß $W_\lambda$ ein Unterverband von $V$ ist. $W_\sigma$ bzw. $W_\delta$ entsteht aus $W$ durch Adjunktion aller möglichen Vereinigungen bzw. Durchschnitte von abzählbar vielen Elementen aus $W$, die in $V$ existieren. Entsprechende Bedeutungen haben $W_{\sigma\delta}$, $W_{\delta\sigma}$, $W_{\sigma\sigma}$, $W_{\delta\delta}$, $W_{\sigma\delta\sigma}$ usw. Es ist $W_{\sigma\sigma} = W_\sigma$, $W_{\delta\delta} \neq W_\delta$.

Es gilt

(3.1). *Satz.* $W_\sigma$ (bzw. $W_\delta$) *ist identisch mit dem Untersystem von $V$, bestehend aus allen Elementen von $V$, die algebraische Limiten von aufsteigenden (bzw. absteigenden) Folgen aus Elementen von $W$ sind.*

*Beweis.* Es sei $a_n \subseteq a_{n+1}$ (bzw. $a_n \supseteq a_{n+1}$) mit $a_n \in W$, $n = 1, 2, \ldots$, und $\lim \operatorname{alg} a_n = a \in V$, dann ist $a = \bigcup a_n$ (bzw. $a = \bigcap a_n$) also $a \in W$ (bzw. $W_\delta$). Es sei umgekehrt $a \in W_\sigma$ (bzw. $W_\delta$), also $a = \bigcup a_n$ (bzw. $= \bigcap a_n$) mit $a \in V$, $a_n \in W$. Wir setzen $s_n = a_1 \cup a_2 \cup \ldots \cup a_n$ ($d_n = a_1 \cap a_2 \cap a_3 \cap \ldots \cap a_n$); dann gilt $s_n \subseteq s_{n+1}$ ($d_n \supseteq d_{n+1}$) und $a = \lim \operatorname{alg} s_n$ (bzw. $= \lim \operatorname{alg} d_n$). Damit ist der Satz bewiesen.

Aus (3.1) und (2.1) folgt:

---

5) HAUPT-AUMANN-PAUC: Differential- und Integralrechnung Bd. I, S. 32, 2. Aufl. Berlin 1948.

6) BIRKHOFF, G.: l. c. 4), § 37.

7) HAHN, H.: l. c. 2) S. 17; auch HAUPT-AUMANN-PAUC, l. c. 5).

(3.2). *Satz.* $W_\sigma$ *bzw.* $W_\delta$ *und allgemein jedes Borelsche Untersystem* $W_{\varrho_1\varrho_2\cdots\varrho_\nu}$ *von* $V$ *über* $W$, *wobei* $\varrho_i = \delta$ oder $= \sigma$, $i = 1, 2, \ldots, \nu$, *ist ein Unterverband von* $V$.

*Anmerkung.* Unter der Voraussetzung der endlichen Distributivität von $V$ kann man mit Hilfe von (2.1) zeigen, daß der topologische Verband $V$ und jeder Unterverband $W$ von $V$ $\sigma$- und $\delta$-distributiv ist, d. h. aus $a = \bigcup a_n$, $b = \bigcup b_n$ mit $a, b, a_n, b_n \in V$ bzw. $W$ folgt $a \cap b = \bigcup\limits_{n,m} (a_n \cap b_m)$ (die $\delta$-Distributivität ist dual zu erklären). Umgekehrt folgt aus der $\sigma$- und $\delta$-Distributivität von $V$, daß $V$ topologisch ist. Es ist also „*endlich distributiv und topologisch*" gleichwertig mit „*$\sigma$- und $\delta$-distributiv*"[8]). Für einen distributiven und topologischen Verband $V$ läßt sich (3.2) direkt beweisen. Wir bemerken ferner, daß ein topologischer Verband nicht immer distributiv zu sein braucht. Als Beispiel betrachten wir die Menge $K$ aller cartesischen Produkte $a_1 \times a_2$ mit $a_1 \in B_1$, $a_2 \in B_2$, wobei $B_1$ und $B_2$ $\sigma$-Boolesche Verbände sind[9]). Durch Definition I § 5[9]) wird $K$ teilweise geordnet. $K$ ist ein $(\sigma, \delta)$-Verband; denn es gilt $\bigcup\limits_k (a_1^{(k)} \times a_2^{(k)}) = \bigcup\limits_k a_1^{(k)} \times \bigcup\limits_k a_2^{(k)}$ bzw. $\bigcap\limits_k (a_1^{(k)} \times a_2^{(k)}) = \bigcap\limits_k a_1^{(k)} \times \bigcap\limits_k a_2^{(k)}$. $K$ ist aber nicht distributiv[10]). $K$ ist topologisch: Es seien nämlich zwei aufsteigende Folgen aus $K$:

$a_1^{(\mu)} \times a_2^{(\mu)} \subseteq a_1^{(\mu+1)} \times a_2^{(\mu+1)}$ bzw. $b_1^{(\mu)} \times b_2^{(\mu)} \subseteq b_1^{(\mu+1)} \times b_2^{(\mu+1)}$, $\mu = 1, 2, \ldots$;

dann ist, wie man leicht bestätigt:

$$\text{(I)}\quad \lim\operatorname{alg}(a_1^{(\mu)} \times a_2^{(\mu)}) \cup \lim\operatorname{alg}(b_1^{(\mu)} \times b_2^{(\mu)}) = \lim\operatorname{alg}\{(a_1^{(\mu)} \times a_2^{(\mu)}) \cup (b_1^{(\mu)} \times b_2^{(\mu)})\}$$

Es gilt aber außerdem:

$$\text{(II)}\quad \lim\operatorname{alg}(a_1^{(\mu)} \times a_2^{(\mu)}) \cap \lim\operatorname{alg}(b_1^{(\mu)} \times b_2^{(\mu)}) = \lim\operatorname{alg}\{(a_1^{(\mu)} \times a^{(\mu)}) \cap (b_1^{(\mu)} \times b_2^{(\mu)})\};$$

denn

$$\begin{aligned}
\lim\operatorname{alg}(a_1^{(\mu)} \times a_2^{(\mu)}) \cap \lim\operatorname{alg}(b_1^{(\mu)} \times b_2^{(\mu)}) &= \bigcup_\mu (a_1^{(\mu)} \times a_2^{(\mu)}) \cap \bigcup_\mu (b_1^{(\mu)} \times b_2^{(\mu)})\\
&= \{\bigcup_\mu a_1^{(\mu)} \times \bigcup_\mu a_2^{(\mu)}\} \cap \{\bigcup_\mu b_1^{(\mu)} \times \bigcup_\mu b_2^{(\mu)}\}\\
&= \{\bigcup_\mu a_1^{(\mu)} \cap \bigcup_\mu b_1^{(\mu)}\} \times \{\bigcup_\mu a_2^{(\mu)} \cap \bigcup b_2^{(\mu)}\}\\
&= \bigcup_\mu (a_1^{(\mu)} \cap b_1^{(\mu)}) \times \bigcup_\mu (a_2^{(\mu)} \cap b_2^{(\mu)})\\
&= \bigcup_\mu \{(a_1^{(\mu)} \cap b_1^{(\mu)}) \times (a_2^{(\mu)} \cap b_2^{(\mu)})\} = \bigcup_\mu \{(a_1^{(\mu)} \times a_2^{(\mu)}) \cap (b_1^{(\mu)} \times b_2^{(\mu)})\}\\
&= \lim\operatorname{alg}\{(a_1^{(\mu)} \times a_2^{(\mu)}) \cap (b_1^{(\mu)} \times b_2^{(\mu)})\}.
\end{aligned}$$

Dual zeigt man, daß (I) und (II) für zwei absteigende Folgen gelten. Daraus folgt aber, daß $K$ topologisch ist.

Wir bemerken endlich, daß Verbände existieren, die nicht topologisch sind[11]).

---

[8]) Birkhoff: l. c. 4), S. 85.

[9]) Kappos, D. A.: Math. Ann. **120**, 43—74 (1947).

[10]) Kappos: l. c. 9), S. 52,

[11]) Birkhoff: l. c. 4). S. 30 Fußnote.

Der in der Einleitung angekündigte Satz lautet nun:

(3.3). Satz. *Es sei $V$ ein bedingter $(\sigma, \delta)$-Verband, der topologisch ist. Ferner sei $W$ irgendein Unterverband von $V$. Dann ist*

$$W_\lambda = W_{\sigma\delta} \cap W_{\delta\sigma}.$$

*Beweis.* Zuerst gilt $W_\lambda \subseteq W_{\sigma\delta} \cap W_{\delta\lambda}$; denn ist $a \in W_\lambda$, so existiert eine Folge $a_1, a_2, \ldots$ aus $W$, so daß $a = \lim\operatorname{alg} a_n = \bigcup_{\mu=1}^{\infty} \bigcap_{\varrho=0}^{\infty} a_{\mu+\varrho} = \bigcap_{\mu=1}^{\infty} \bigcup_{\varrho=0}^{\infty} a_{\mu+\varrho}$, also $a \in W_{\delta\sigma}$ und $a \in W_{\sigma\delta}$. Es bleibt daher zu zeigen: $W_{\sigma\delta} \cap W_{\delta\sigma} \subseteq W_\lambda$, d. h. aus $a \in W_{\delta\sigma} \cap W_{\sigma\delta}$ folgt $a \in W_\lambda$.

Aus $a \in W_{\delta\sigma} \cap W_{\sigma\delta}$ folgt aber:

(3.4) $$a = \bigcup_{i=1}^{\infty} c_i = \bigcap_{i=1}^{\infty} b_i \text{ mit } c_i \in W_\delta,\ b_i \in W_\sigma,\ i = 1, 2, \ldots;$$

(3.5) $$b_i = \bigcup_{k=1}^{\infty} b_{ik},\ c_i = \bigcap_{k=1}^{\infty} c_{ik} \text{ mit } b_{ik},\ c_{ik} \in W,\ i, k = 1, 2, \ldots$$

Wegen Satz (3.1) und (3.2) kann angenommen werden:

(3.6) $$b_1 \supseteq b_2 \supseteq \ldots \quad \text{bzw.} \quad c_1 \subseteq c_2 \subseteq \ldots$$

und

(3.7) $$b_{i1} \subseteq b_{i2} \subseteq \ldots \quad \text{bzw.} \quad c_{i1} \supseteq c_{i2} \supseteq \ldots,\ i = 1, 2, \ldots$$

Außerdem kann o.B.d.A. vorausgesetzt werden, daß

(3.8) $$b_{1k} \supseteq b_{2k} \supseteq b_{3k} \supseteq \ldots \quad \text{bzw.} \quad c_{1k} \subseteq c_{2k} \subseteq c_{3k} \subseteq \ldots,\ k = 1, 2, \ldots$$

Ist nämlich (3.8) nicht erfüllt, so ersetzen wir $b_{\varrho k}$ durch

$$b'_{\varrho k} = b_{1k} \cap \cdots \cap b_{\varrho k} \quad \text{bzw.} \quad c_{\varrho k} \text{ durch } c'_{\varrho k} = c_{1k} \cup c_{2k} \cup \cdots \cup c_{\varrho k}$$

und haben $b'_{\varrho k} \subseteq b'_{\varrho, k+1}$, $c'_{\varrho k} \supseteq c'_{\varrho, k+1}$ sowie

$$\lim\operatorname{alg} b'_{\varrho k} = \bigcup_{k=1}^{\infty} b'_{\varrho k} = \lim\operatorname{alg} b_{1k} \cap \lim\operatorname{alg} b_{2k} \cap \ldots \cap \lim\operatorname{alg} b_{\varrho k} =$$
$$= b_1 \cap b_2 \cap \ldots \cap b_\varrho = b_\varrho$$

und entsprechend $\bigcap_{k=1}^{\infty} c'_{\varrho k} = c_\varrho$.

Unter der Voraussetzung, daß (3.6), (3.7) und (3.8) gelten, bilden wir jetzt die Elemente

(3.9) $$a_n = (c_{1n} \cap b_{1n}) \cup (c_{2n} \cap b_{2n}) \cup \cdots \cup (c_{nn} \cap b_{nn}) \text{ für } n = 1, 2, \ldots.$$

Es ist $a_n \in W$ für alle $n = 1, 2, \ldots$. Wir werden zeigen, daß $\lim\operatorname{alg} a_n$ existiert und gleich $a$ ist, d. h. $a \in W_\lambda$.

Es ist nämlich

$$\bigcup_{n=k}^{\infty} (c_k \cap b_{kn}) = \lim_{n\to\infty}\operatorname{alg} (c_k \cap b_{kn}) = c_k \cap b_k = c_k \text{ für } k = 1, 2, \ldots,$$

bzw.

$$\bigcap_{n=k}^{\infty} (b_k \cup c_{kn}) = \lim\operatorname{alg} (b_k \cup c_{kn}) = b_k \cup c_k = b_k \text{ für } k = 1, 2, \ldots,$$

also

$$a = \bigcup_{k=1}^{\infty} c_k = \bigcup_{k=1}^{\infty} \bigcup_{n=k}^{\infty} (c_k \cap b_{kn}) = \begin{array}{r} (c_1 \cap b_{11}) \cup (c_1 \cap b_{12}) \cup (c_1 \cap b_{13}) \cup \ldots \\ \cup (c_2 \cap b_{22}) \cup (c_2 \cap b_{23}) \cup \ldots \\ \cup (c_3 \cap b_{33}) \cup \ldots \end{array}$$

weil das kommutative bzw. assoziative Gesetz von $\cup$ für beliebig viele Glieder gilt, so ist

$$a = \bigcup_{n=1}^{\infty} \{(c_1 \cap b_{1n}) \cup (c_2 \cap b_{2n}) \cup \cdots \cup (c_n \cap b_{nn})\}$$

oder

$$(3.10) \qquad a = \bigcup_{n=1}^{\infty} a_n \text{ mit } a'_n = (c_1 \cap b_{1n}) \cup (c_2 \cap b_{2n}) \cup \cdots \cup (c_n \cap b_{nn}).$$

Entsprechend können wir

$$(3.11) \qquad a = \bigcap_{n=1}^{\infty} a''_n \text{ mit } a''_n = (b_1 \cup c_{1n}) \cap (b_2 \cup c_{2n}) \cap \cdots \cap (b_n \cup c_{nn})$$

setzen. Es ist aber $a'_n \subseteq a'_{n+1}$ bzw. $a''_n \supseteq a''_{n+1}$ für $n = 1, 2, \ldots$, so daß wegen (3.10) und (3.11) gilt

$$(3.12) \qquad a = \lim \operatorname{alg} a'_n = \lim \operatorname{alg} a''_n.$$

Nun ist einerseits:

$$(\alpha) \qquad \begin{aligned} a'_n &= (c_1 \cap b_{1n}) \cup (c_2 \cap b_{2n}) \cup \cdots \cup (c_n \cap b_{nn}) \subseteq \\ &\subseteq (c_{1n} \cap b_{1n}) \cup (c_{2n} \cap b_{2n}) \cdots \cup (c_{nn} \cap b_{nn}) = a_n, \end{aligned}$$

weil $c_k \subseteq c_{kn}$ für alle $k = 1, 2, \ldots, n$ wegen (3.5) gilt, und andererseits:

$$(\beta) \qquad \begin{aligned} a_n &= (c_{1n} \cap b_{1n}) \cup (c_{2n} \cap b_{2n}) \cup \cdots \cup (c_{nn} \cap b_{nn}) \subseteq \\ &\subseteq (b_1 \cup c_{1n}) \cap (b_2 \cup c_{2n}) \cap \cdots \cap (b_n \cup c_{nn}) = a''_n. \end{aligned}$$

Dies ergibt sich aus $b_{\mu n} \cap c_{\mu n} \subseteq b_k \cup c_{kn}$ für alle $k = 1, 2, \ldots, n$ und $\mu = 1, 2, \ldots, n$. Letzteres folgt so: Für $\mu = k$ ist $c_{\mu n} = c_{kn}$ und $b_k \subseteq b_k$ wegen (3.5); für $\mu > k$ ist $c_{\mu n} \cap b_{\mu n} \subseteq b_{\mu n} \subseteq b_{kn} \subseteq b_k \subseteq b_k \cup c_{kn}$ wegen (3.8); und für $\mu < k$ ist $c_{\mu n} \cap b_{\mu n} \subseteq c_{\mu n} \subseteq c_{kn} \subseteq b_k \cup c_{kn}$ wegen (3.8).
Aus ($\alpha$) und ($\beta$) folgt $a'_n \subseteq a_n \subseteq a''_n$ und da $\lim \operatorname{alg} a'_n = \lim \operatorname{alg} a''_n = a$, so existiert $\lim \operatorname{alg} a_n$ und ist auch gleich $a$ wegen (1.1). Damit ist der Satz bewiesen.

*(Eingegangen am 4. Juni 1949.)*

Math. Annalen, Bd. 122, S. 6—13 (1950).

# Konvergenzbetrachtung zum Abbildungsverfahren von Theodorsen-Garrick[1]).

Von

HANS WITTICH in Karlsruhe.

1. Ein einfach zusammenhängendes Gebiet $G_w$ in der $w$-Ebene mit dem Rand $\Gamma_w$ sei bezüglich $w = 0$ sternförmig im strengen Sinne. Zwischen den Randpunkten $w = Pe^{i\Theta}$ und dem Polarwinkel $\theta$, $0 \leq \theta < 2\pi$, besteht also eine eindeutige Zuordnung $P = P(\theta)$. Es gibt genau eine Funktion $w(z) = z + a_2 z^2 + \cdots$, die $|z| < R$ schlicht und konform in $G_w$ abbildet, wobei der Abbildungsradius $R$ zufolge der Normierung durch $G_w$ eindeutig bestimmt ist. Da beide Gebiete, $|z| < R$ und $G_w$, nur erreichbare Randpunkte besitzen, bezieht $w = w(z)$ die Ränder der Gebiete umkehrbar eindeutig und stetig aufeinander. Die Funktion $\frac{w(z)}{z} = 1 + a_2 z + \cdots$ ist in $|z| < R$ eindeutig regulär analytisch und $\neq 0$, so daß also nach Fixierung eines Zweiges die Funktion

$$h(z) = \log \frac{w(z)}{z} = A_1 z + \cdots, \log\left(\lim_{z \to 0} \frac{w(z)}{z}\right) = 0, \tag{1}$$

in $|z| < R$ eindeutig analytisch ist. Wegen $\frac{w(z)}{z} \neq 0, \infty$ auf $|z| \leq R$ gilt nach der Poisson-Jensenschen Formel die Darstellung

$$h(z) = \frac{1}{2\pi} \int_0^{2\pi} \log \left| \frac{w(Re^{i\alpha})}{Re^{i\alpha}} \right| \frac{Re^{i\alpha} + z}{Re^{i\alpha} - z} d\alpha, z = re^{i\varphi}, \tag{2}$$

mit $\log\left|\frac{w(Re^{i\alpha})}{Re^{i\alpha}}\right| = \log \frac{P(\alpha)}{R}$. Unter der Voraussetzung, daß die Funktion $\log P(\alpha) = H(\alpha)$ einer Hölderbedingung genügt, — das besagt, daß $H(\alpha)$ eine Ungleichheit von der Form $|H(\alpha + \delta) - H(\alpha)| \leq A\,|\delta|^p$, $A$ konstant und $0 < p < 1$, erfüllt — erhält man durch Grenzübergang $z \to \zeta = Re^{i\varphi}$

$$\theta(\varphi) = \varphi - \frac{1}{2\pi} \int_0^{2\pi} \log \frac{P(\alpha)}{R} \operatorname{ctg} \frac{\alpha - \varphi}{2} d\alpha; \tag{3}$$

oder mit $g(\varphi) = \theta(\varphi) - \varphi$ und $\log P(\theta(\varphi)) = f(\theta(\varphi)) = f(\varphi + g(\varphi))$

$$g(\varphi) = -\frac{1}{2\pi} \int_0^{2\pi} f(\alpha + g(\alpha)) \operatorname{ctg} \frac{\alpha - \varphi}{2} d\alpha. \tag{J}$$

---

[1]) KARMAN, TH. v., und E. TREFFTZ: Potentialströmung um gegebene Tragflächenquerschnitte. Z. Flugtechn. u. Motorl. **9** (1918). — THEODORSEN, TH., a. J. E. GARRICK, General potential theory of arbitrary wing sections. National advisory commitee aeronautics. Rep. **1933**, 452.

Der wesentliche Inhalt dieser Arbeit wurde auf der Jahrestagung der dtsch. math. Ver. 1943 in Würzburg vorgetragen. Die Anfang 1944 beim Jber. der dtsch. math. Ver. eingereichte Ausarbeitung wurde der Kriegsereignisse wegen nicht mehr abgedruckt.

Dieser Integralgleichung muß die Funktion $g(\varphi)$ genügen. Aus $\int_0^{2\pi} h(Re^{i\alpha})\,d\alpha = 0$ folgt

$$\log R = \frac{1}{2\pi}\int_0^{2\pi} f(\alpha + g(\alpha))\,d\alpha \text{ und } \int_0^{2\pi} g(\alpha)\,d\alpha = 0. \tag{4}$$

2. Zur Konstruktion der Abbildungsfunktion $w = w(z)$ kann man bei gegebenem $G_w$, also bekanntem $f(\theta)$, von der Integralgleichung ($J$) ausgehen, eine Lösung $g(\varphi)$ bestimmen und daraus nach (1), (2), und (4) schließlich die gesuchte Abbildungsfunktion aufbauen. Danach ist zunächst die Frage nach Existenz und Eindeutigkeit einer Lösung von ($J$) zu untersuchen.

Zum Existenznachweis kann man sukzessive Approximationen anwenden:

$$\begin{aligned} g_n(\varphi) &= -\frac{1}{2\pi}\int_0^{2\pi} f(\alpha + g_{n-1}(\alpha))\operatorname{ctg}\frac{\alpha-\varphi}{2}\,d\alpha \\ &= -\frac{1}{2\pi}\int_0^{2\pi} f_{n-1}(\alpha)\operatorname{ctg}\frac{\alpha-\varphi}{2}\,d\alpha, \end{aligned} \tag{5}$$

$n = 1, 2, \cdots$. Für $g_0(\varphi)$ wähle man eine $2\pi$-periodische Funktion, die einer Hölderbedingung genügen möge. $f(\theta)$ sei auf $\langle 0, 2\pi\rangle$ stetig differenzierbar, $|f'(\theta)| \le M$. Dann gilt, wie man sofort bestätigt,

$$|f(\varphi + g_0(\varphi)) - f(\varphi_0 + g_0(\varphi_0))| = |f_0(\varphi) - f_0(\varphi_0)| < G_0\,|\varphi - \varphi_0|^p.$$

Nach einem Satz von FATOU[2]) genügt dann auch die Funktion $g_1(\varphi)$ einer Hölderbedingung $|g_1(\varphi) - g_1(\varphi_0)| < D_1(p)\,|\varphi - \varphi_0|^p$. Vollständige Induktion zeigt, daß alle Integrale $\int_0^{2\pi} f(\alpha + g_n(\alpha))\operatorname{ctg}\frac{\alpha-\varphi}{2}\,d\alpha$ existieren und für alle Funktionen $g_n(\varphi)$, $f_n(\varphi) = f(\varphi + g_n(\varphi))$ gilt

$$\begin{aligned} |f_n(\varphi) - f_n(\varphi_0)| &< G_n\,|\varphi - \varphi_0|^p \\ |g_n(\varphi) - g_n(\varphi_0)| &< D_n\,|\varphi - \varphi_0|^p, \quad 0 < p < 1. \end{aligned}$$

Durch (5) ist also eine Funktionenfolge $g_n(\varphi)$ erklärt, deren Konvergenz zu untersuchen ist.

Die Integralgleichung sei von der etwas allgemeineren Form

$$g(\varphi) = F(\varphi) - \frac{1}{2\pi}\int_0^{2\pi} f(\alpha + g(\alpha))\operatorname{ctg}\frac{\alpha-\varphi}{2}\,d\alpha, \tag{6}$$

und folgende Voraussetzungen seien erfüllt:

1. $F(\varphi)$ sei $2\pi$-periodisch, auf $\langle 0, 2\pi\rangle$ differenzierbar und $F'(\varphi)$ genüge einer Hölderbedingung der Ordnung $p$.

2. Die $2\pi$-periodische Funktion $f(\theta)$ sei auf $\langle 0, 2\pi\rangle$ differenzierbar, und $f'(\theta)$ genüge dort einer Hölderbedingung der Ordnung $p$. Weiterhin gelte $M < 1$.

3. Die Ausgangsnäherung $g_0(\varphi)$ sei auf $\langle 0, 2\pi\rangle$ differenzierbar, und $g_0'(\varphi)$ genüge einer Hölderbedingung.

---

[2]) FATOU, P.: Séries trigonométriques et series de TAYLOR. Acta math. 30.

Die Voraussetzung über $g_0(\varphi)$ bedeutet kaum eine Einschränkung, da man bei der praktischen Rechnung meistens $g_0(\varphi) \equiv 0$ setzt. Unter diesen Voraussetzungen hat die Folge $g_n(\varphi)$ die folgenden Eigenschaften:

a) Die $g_n(\varphi)$ haben gleichmäßig beschränkte Normen

$$N g_n = \int_0^{2\pi} g_n^2(\varphi)\, d\varphi \leq K_1 < \infty. \tag{7}$$

b) Es gilt bei beliebigem $\varepsilon > 0$ für alle $m \geq 1$ und alle $n > n(\varepsilon)$

$$N(g_{n+m} - g_n) < \varepsilon. \tag{8}$$

c) Die $g_n(\varphi)$ bilden eine gleichgradig stetige Funktionenmenge, d. h.

$$|g_n(\varphi) - g_n(\varphi_0)| < \varepsilon \text{ für alle } n, \text{ falls nur } |\varphi - \varphi_0| < \delta = \delta(\varepsilon, \varphi_0). \tag{9}$$

Aus a), b) und c) folgt nach einem Konvergenzsatz über gleichgradig stetige Funktionen, daß auf $\langle 0, 2\pi \rangle$ gleichmäßig $\lim\limits_{n\to\infty} g_n(\varphi) = g(\varphi)$ gilt.

*Beweis zu a).* Unter den angegebenen Voraussetzungen gilt nach der Vollständigkeitsrelation

$$N(g_n - F) \leq N f_{n-1} = \int_0^{2\pi} f^2(\alpha + g_{n-1}(\alpha))\, d\alpha \leq K^2\, 2\pi$$

und wegen $N(g_n - F) \geq \left(\sqrt{N g_n} - \sqrt{N F}\right)^2$

$$\left(\sqrt{N g_n} - \sqrt{N F}\right)^2 \leq K^2\, 2\pi \text{ oder } \sqrt{N g_n} \leq \sqrt{N F} + K\sqrt{2\pi}\text{, also}$$

$$N g_n \leqq K_1 < \infty.$$

*Beweis zu b).* Aus

$$\int_0^{2\pi} (g_n - g_{n-1})^2\, d\varphi \leq \int_0^{2\pi} (f_{n-1} - f_{n-2})^2\, d\varphi \leq M^2 \int_0^{2\pi} (g_{n-1} - g_{n-2})^2\, d\varphi,$$

also $N(g_n - g_{n-1}) \leq M^2 N(g_{n-1} - g_{n-2})$, folgt durch wiederholte Anwendung

$$N(g_n - g_{n-1}) < M^{2(n-1)} N(g_1 - g_0) = K_2 M^{2(n-1)}. \tag{10}$$

Nun gilt wegen $(g_{n+m} - g_n)^2 = \left(\sum\limits_{\mu=1}^{m} (g_{n+\mu} - g_{n+\mu-1})\right)^2 = \left(\sum\limits_{\mu=1}^{m} \Gamma_\mu\right)^2$

$$= \Gamma^2 + \cdots + \Gamma_m^2 + 2\,\Gamma_1\Gamma_2 + \cdots 2\,\Gamma_{m-1}\Gamma_m \text{ und}$$

$$\left(\int_0^{2\pi} \Gamma_\varkappa(\varphi)\,\Gamma_\lambda(\varphi)\, d\varphi\right)^2 \leq \int_0^{2\pi} \Gamma_\varkappa^2(\varphi)\, d\varphi \int_0^{2\pi} \Gamma_\lambda^2(\varphi)\, d\varphi$$

$$N(g_{n+m} - g_n) \leq N\Gamma_1 + \cdots + N\Gamma_m + 2\sqrt{N\Gamma_1}\cdot\sqrt{N\Gamma_2} + \cdots + 2\sqrt{N\Gamma_{m-1}}\,\sqrt{N\Gamma_m}$$

$$= \left(\sum_{\mu=1}^{m} \sqrt{N\Gamma_\mu}\right)^2\text{, also}$$

$$N(g_{n+m} - g_n) \leq \left(\sum_{\mu=1}^{m} \sqrt{N(g_{n+\mu} - g_{n+\mu-1})}\right)^2.$$

Daraus erhält man wegen (10) und $M < 1$

$$N(g_{n+m} - g_n) \leq K_2 M^{2n} (\sum_{\mu=1}^{m} M^{\mu-1})^2 \leq K_2 \frac{M^{2n}}{(1-M)^2} = K_3 M^{2n}, \text{ also (8)}.$$

*Beweis zu c).* Es existiere $g_n'(\varphi)$ auf $\langle 0, 2\pi \rangle$ und genüge einer Hölderbedingung. Da aus $\frac{d f_n}{d\varphi} = \frac{d f}{d\Theta} \frac{d}{d\varphi}(\varphi + g_n(\varphi))$ die Beziehung $|f_n'(\varphi) - f_n'(\varphi_0)| < K_4 |\varphi - \varphi_0|^p$ folgt, wobei $K_4$ von $n$ abhängen kann, genügt auch $f_n'(\varphi)$ einer Hölderbedingung. Wegen $g_{n+1}'(\varphi) = F'(\varphi) - \frac{1}{2\pi} \times \times \int_0^{2\pi} f_n'(\alpha) \operatorname{ctg} \frac{\alpha - \varphi}{2} d\alpha$ folgt aus dem Satz von FATOU, daß die Funktion $g_{n+1}'(\varphi)$ eine Hölderbedingung erfüllt. Da nach 3. $g_0'(\varphi)$ eine solche erfüllt, gilt für alle $n$

$$(11) \qquad |g_n'(\varphi) - g_n'(\varphi_0)| < D_n'(p) |\varphi - \varphi_0|^p.$$

Nach der Vollständigkeitsrelation gilt

$$N(g_{n+1}' - F') \leq N f_n' \leq M^2 \int_0^{2\pi} (1 + g_n'^2) d\varphi, \text{ also}$$

$$(12) \qquad N(g_{n+1}' - F') \leq M^2 (2\pi + N g_n') < M^2 \left(\sqrt{2\pi} + \sqrt{N g_n'}\right)^2.$$

Daraus folgt wegen $N(g_{n+1}' - F') \geq \left(\sqrt{N g_{n+1}'} - \sqrt{N F'}\right)^2$

$$\sqrt{N g_{n+1}'} \leq \sqrt{N F'} + M\left(\sqrt{2\pi} + \sqrt{N g_n'}\right) = M\left(c + \sqrt{N g_n'}\right)$$

mit $c = \sqrt{2\pi} + \frac{\sqrt{N F'}}{M}$, gültig für $n = 0, 1, 2, \ldots$. Wiederholte Anwendung dieser Formel ergibt

$$\sqrt{N g_n'} \leq M c + M \sqrt{N g_{n-1}'} \leq \cdots \leq c(M + M^2 + \cdots + M^n) + M^n \sqrt{N g_0'}$$

$$\sqrt{N g_n} < \frac{c}{1 - M} + \sqrt{N g_0} < \sqrt{K_5}$$

oder

$$N g_n' < K_5.$$

Aus der Tatsache, daß die Normen von $g_n'(\varphi)$ gleichmäßig beschränkt sind, folgt die Behauptung c) in bekannter Weise so:

$$|g_n(\varphi) - g_n(\varphi_0)|^2 = \left(\int_{\varphi_0}^{\varphi} g_n'(\alpha) d\alpha\right)^2 \leq |\varphi - \varphi_0| \int_0^{2\pi} g_n'^2(\alpha) d\alpha \leq K_5 |\varphi - \varphi_0|$$

oder

$$(13) \qquad |g_n(\varphi) - g_n(\varphi_0)| \leq \sqrt{K_5} |\varphi - \varphi_0|^{1/2}, \text{ also (9)}^{3)}.$$

---

[3]) Die gleichgradige Stetigkeit der Funktionen $g_n(\varphi)$ ist auch bewiesen, wenn man zeigt, daß in $|g_n(\varphi) - g_n(\varphi_0)| < D_n |\varphi - \varphi_0|^p$ die $D_n$ für alle $n$ der Bedingung $D_n \leq \leq D(p)$ genügen, wobei $D(p)$ eine nur von $p$ abhängige endliche Konstante ist. Durch Verfeinerung der Abschätzungen von FATOU kommt man auf diesem Wege ebenfalls zu hinreichenden Konvergenzbedingungen, die allerdings schlechter sind als die hier angegebene, wenn man die zur Verfügung stehenden Abschätzungen benützt. Es genügt dann zu fordern, daß die Ausgangsnäherung $g_0(\varphi)$ einer Hölderbedingung der Ordnung $p$ genügt und $f(\Theta)$ auf $\langle 0, 2\pi \rangle$ stetig differenzierbar ist. Die Voraussetzungen über $f(\Theta)$ lassen sich noch etwas lockern, worauf aber hier nicht eingegangen werden soll.

Danach gibt es eine stetige Funktion $g(\varphi)$ derart, daß auf $0 \leq \varphi \leq 2\pi$ gleichmäßig $\lim_{n\to\infty} g_n(\varphi) = g(\varphi)$ gilt. Aus (13) folgt bei $n \to \infty$, daß $g(\varphi)$ sicher einer Hölderbedingung der Ordnung $p = \frac{1}{2}$ genügt. Daher existiert die Funktion $G(\varphi)$

$$G(\varphi) = F(\varphi) - \frac{1}{2\pi}\int_0^{2\pi} f(\alpha + g(\alpha)) \operatorname{ctg}\frac{\alpha - \varphi}{2}\, d\alpha$$

und erfüllt eine Hölderbedingung. Aus $N(G - g_n) \leq M^2 N(g - g_{n-1}) < 2\pi M^2 \varepsilon^2$ für alle $n \geq n(\varepsilon)$ folgt $N(G - g) = 0$, also wegen der Stetigkeit des Integranden $G(\varphi) \equiv g(\varphi)$; $g(\varphi)$ ist daher tatsächlich Lösung der Integralgleichung (6). Unter den erwähnten Bedingungen ist $g(\varphi)$ auch die einzige stetige Lösung. Ist nämlich $\bar{g}(\varphi)$ eine zweite stetige Lösung, so gilt

$$N(g - \bar{g}) \leq M^2 N(g - \bar{g}), \text{ also } g(\varphi) \equiv \bar{g}(\varphi).$$

Unter den Voraussetzungen 1., 2., und 3. ist die nichtlineare Integralgleichung (6) eindeutig lösbar mit stetigem $g(\varphi)$, und die Lösung $g(\varphi)$ läßt sich durch sukzessive Approximationen berechnen:

$$g_n(\varphi) = F(\varphi) - \frac{1}{2\pi}\int_0^{2\pi} f(\alpha + g_{n-1}(\alpha)) \operatorname{ctg}\frac{\alpha - \varphi}{2}\, d\alpha, \quad n = 0, 1, \cdots.$$

3. Es sei nun ein bezüglich $w = 0$ sternförmiges Gebiet $G_w$ der $w$-Ebene mit der Randkurve $\Gamma_w : P = P(\theta)$, $0 \leq \theta \leq 2\pi$, gegeben. Die Funktion $\log P(\theta) = f(\theta)$ sei differenzierbar und $f'(\theta)$ genüge einer $H$.-Bedingung der Ordnung $p$, $0 < p < 1$. Weiter gelte noch $\operatorname{Max}|f'(\theta)| = M < 1$. Gesucht ist die Funktion $w = w(z)$, die $|z| < R$ schlicht, normiert und konform in das vorgegebene Gebiet $G_w$ abbildet.

Zur Konstruktion setze man in (6) $F(\varphi) \equiv 0$ und wähle eine zulässige Ausgangsnäherung $g_0(\varphi)$, z. B. $g_0(\varphi) \equiv 0$. Da die Voraussetzungen 1. bis 3. erfüllt sind, hat (6) eine eindeutig bestimmte stetige Lösung $g(\varphi)$. Aus $g(\varphi)$ bestimmt sich nach (4) der Abbildungsradius $R$. Aus $g(\varphi)$ und der gegebenen Funktion $f(\theta)$ wird die eindeutige und stetige Funktion $f(\varphi + g(\varphi)) = f(\theta(\varphi)) = \Phi(\varphi)$ gebildet. Eine in $|z| < R$ eindeutige und regulär analytische Funktion $h(z)$ mit den Randwerten $\Phi(\varphi) - \log R$ des Realteils ist bis auf eine Konstante bestimmt und schreibt sich in der Form

$$h(z) = iC + \frac{1}{2\pi}\int_0^{2\pi} (\Phi(\alpha) - \log R)\frac{R e^{i\alpha} + r e^{i\varphi}}{R e^{i\alpha} - r e^{i\varphi}}\, d\alpha, \; r < R. \tag{14}$$

$w = w(z) = z\, e^{h(z)}$ ist in $|z| < R$ eindeutig und regulär analytisch. Aus der Forderung $\lim_{z\to 0} \frac{w(z)}{z} = 1$ folgt $C = 0$. Zufolge der über $G_w$ gemachten Voraussetzungen ist, wenn man $\zeta = R e^{i\varphi}$ setzt, $\lim_{z\to\zeta} w(z) = R\, e^{i\varphi} \times \times e^{\Phi(\varphi) - \log R + i g(\varphi)} = P e^{i\theta}$. Die Randwerte $w(\zeta)$ der Funktion $w = w(z)$ liegen also auf $\Gamma_w$.

Der Wert $w = 0$ wird von der Funktion $w(z) = z\, e^{h(z)}$ im Punkte $z = 0$ genau einmal angenommen und nur dort. Danach gibt es einen Kreis

$K_\varrho : |w| \leq \varrho$, so daß jeder Wert $w_0$ auf $K_\varrho$ von $w(z)$ in $|z| < R$ genau einmal angenommen wird. Davon ausgehend zeigt man, etwa durch einen Kreiskettenschluß, daß für jeden beliebigen Punkt $w_0$ aus $G_w$ gilt: $\Delta_{|z|=R} \arg (w - w_0) = 2\pi$, d. h. wird der Kreis $|z| = R$ einmal in positivem Sinne durchlaufen, so ändert sich das Argument von $w(z) - w_0$ um $2\pi$. Das bedeutet also, daß die Funktion $w = w(z)$ jeden Wert $w_0$ aus $G_w$ in $|z| < R$ genau einmal annimmt. Weiter folgt aus dem Argumentprinzip, daß $w(z)$ in $|z| < R$ nur Werte aus $G_w$ annehmen kann. Diese Eigenschaften der Funktion $w(z) = z\, e^{h(z)}$ besagen aber, daß von ihr die geforderte Abbildung des Kreises $|z| < R$ in $G_w$ geleistet wird. Daß es auch die einzige Funktion dieser Art ist, folgt in bekannter Weise.

Da beide Gebiete nur erreichbare Randpunkte haben, ist die Abbildungsfunktion auch noch auf den Rändern stetig und umkehrbar eindeutig. Wenn $z$ den Kreis $|z| = R$ einmal in mathematisch positivem Sinne durchläuft, wird vom Bildpunkt die Kurve $\Gamma_w$ auch einmal in mathematisch positivem Sinne durchlaufen, so daß also $\theta = \theta(\varphi)$ mit $\varphi$ monoton wächst[4]).

4. Wendet man auf die in $|z| < R$ harmonischen Funktionen $\Re\, h(z) = \log \left|\frac{w(z)}{z}\right|$ und $\Im\, h(z) = \arg \frac{w(z)}{z}$ das Maximumprinzip an, so ergibt sich

$$\operatorname{Min} \frac{P(\theta)}{R} \leq \left|\frac{w(z)}{z}\right| \leq \operatorname{Max} \frac{P(\theta)}{R}$$

oder

$$|z| \operatorname{Min} \frac{P(\theta)}{R} \leq |w(z)| \leq |z| \cdot \operatorname{Max} \frac{P(\theta)}{R}$$

und $\arg z + \operatorname{Min} g(\varphi) \leq \arg w(z) \leq \arg z + \operatorname{Max} g(\varphi)$.
Verläuft die Randkurve $\Gamma_w$ des Gebietes $G_w$ auf dem Kreisring

$$\varrho(1 - \varepsilon) \leq |w| \leq \varrho(1 + \varepsilon),\ 0 \leq \varepsilon < 1,$$

so erhält man

$$|z| \frac{\varrho}{R}(1 - \varepsilon) \leq |w(z)| \leq |z| \cdot \frac{\varrho}{R}(1 + \varepsilon)$$

oder wegen $\varrho(1 - \varepsilon) \leq R \leq \varrho(1 + \varepsilon)$ (nach (4))

$$|z| \frac{1 - \varepsilon}{1 + \varepsilon} \leq |w(z)| \leq |z| \frac{1 + \varepsilon}{1 - \varepsilon}.$$

Bei kleinem $\varepsilon$ weicht für alle $z$ auf $|z| \leq R$ die Größe $\left|\frac{w(z)}{z}\right|$ wenig von 1 und $\arg \frac{w(z)}{z}$ wenig von 0 ab. Diese plausible Aussage läßt sich nach L. BIEBERBACH[5]) noch etwas präzisieren. Aus der Abschätzung für $P(\theta)$ folgt für die Länge $L$ von $\Gamma_w$

$$\varrho(1 - \varepsilon) < \frac{L}{2\pi} < \varrho(1 + \varepsilon).$$

[4]) Der Konvergenzbeweis läßt sich etwas vereinfachen, wenn man die Existenz der Funktion $g(\varphi)$ oder, was gleichbedeutend ist, die der Abbildungsfunktion $w = w(z)$ annimmt. Von der dabei erforderlichen Anwendung des Riemannschen Abbildungssatzes wurde hier absichtlich abgesehen.

[5]) L. BIEBERBACH: Über die konforme Kreisabbildung nahezu kreisförmiger Bereiche. Sitzgsber. preuß. Akad. Wiss. 1924.

Mit $z = R\zeta$ erhält man $W(\zeta) = w(R\zeta) = R\zeta + \cdots$ und $W'(\zeta) = \frac{dW}{d\zeta} = R + \cdots$, so daß also nach Fixierung eines Zweiges die Funktion $\sqrt{W'(\zeta)} = c_1 + c_2\zeta + \cdots = \sqrt{R} + \cdots$ in $|\zeta| < 1$ eindeutig und regulär analytisch ist. Für $L$ findet man

$$L = 2\pi \sum_{\mu=1}^{\infty} |c_\mu|^2.$$

Aus $|W(\zeta) - R\zeta|^2 \leq \int_0^{|\zeta|} \left|\sqrt{W'(\zeta)} - \sqrt{R}\right|^2 dr \int_0^{|\zeta|} \left|\sqrt{W'(\zeta)} + \sqrt{R}\right|^2 dr$ folgt auf dem von L. Bieberbach angegebenen Wege

$$|W(\zeta) - R\zeta|^2 \leq \pi\left(\frac{L}{2\pi} - R\right)\pi\left(\frac{L}{2\pi} + 3R\right) \leq \pi^2 8 \varrho^2 \varepsilon (1 + \varepsilon)$$

oder

$$|w(z) - z| \leq 2\pi\varrho\sqrt{2\varepsilon(1+\varepsilon)} < 4\pi\varrho\sqrt{\varepsilon}.$$

Diese Abschätzung läßt sich, falls man mehr über den Verlauf der Funktion $P(\theta)$ weiß, in manchen Fällen auf dem Wege über (14) verbessern.

5. Ist $G_w$ das Innere einer Ellipse mit der Gleichung $u = a\cos\theta$, $v = b\sin\theta$ $(b < a)$, so gilt $P(\theta) = a\sqrt{1 - k^2\sin^2\theta}$, $k^2 = 1 - \left(\frac{b}{a}\right)^2$. Die Berechnung der Abbildungsfunktion $w(z)$, die $|z| < R$ schlicht normiert und konform in $G_w$ abbildet, ist nach der Methode der sukzessiven Approximationen sicher möglich, wenn auf $0 \leq \theta \leq 2\pi$ $\left|\frac{df}{d\Theta}\right| = \left|-\frac{k^2\sin\Theta\cos\Theta}{1 - k^2\sin^2\Theta}\right| < 1$ ist. Aus dieser Bedingung erhält man $k^2 < 2\sqrt{2} - 2$ oder

$$0{,}41\ldots = \sqrt{3 - 2\sqrt{2}} < \frac{b}{a} < 1.$$

Nach H. A. Schwarz läßt sich diese Abbildungsaufgabe mittels elliptischer Funktionen in expliziter Form lösen, und man erhält für die Funktion $\varphi = \varphi(\theta)$ eine gut konvergierende Reihenentwicklung, so daß sich diese Beispiele gut zu Betrachtungen über die Güte des numerischen Verfahrens eignen.

Bei den Anwendungen der konformen Abbildung in der Strömungslehre[6]) interessiert man sich in erster Linie für $\frac{dz}{dw}$ auf $\Gamma_w$. Aus den Voraussetzungen in 2. folgt, daß in der für $\Gamma_w$ gültigen Parameterdarstellung $u = u(s), v = v(s)$ $(s =$ Bogenlänge) die Funktionen $\frac{du}{ds}, \frac{dv}{ds}$ eine $H$.-Bedingung erfüllen. Daher ist $\frac{dz}{dw}$ noch auf $\Gamma_w$ stetig und $\neq 0$, also auf $G_w + \Gamma_w$ $\left|\frac{dz}{dw}\right| > 0$. Durch vorbereitende Hilfsabbildungen erzwingt man die Voraussetzungen für die Anwendbarkeit des Verfahrens.

Bei der wirklichen Berechnung von $w(z)$ für $|z| < R$ wird man nicht auf (14) zurückgreifen, sondern von der Fourierentwicklung der Funktion $g(\varphi) = \theta(\varphi) - \varphi$ ausgehen:

[6]) Betz, A.: Konforme Abbildung, vgl. S. 259—264. Springer-Verlag 1948.

$$\theta(\varphi) - \varphi = \sum_{n=1}^{\infty} (-B_n \cos n\varphi + A_n \sin n\varphi) = \sum_{n=1}^{\infty} (-b_n R^n \cos n\varphi + a_n R^n \sin n\varphi).$$

Die konjugierte Reihe $\sum_{n=1}^{\infty} (A_n \cos n\varphi + B_n \sin n\varphi)$ konvergiert auf $0 \leq \varphi \leq \leq 2\pi$ gleichmäßig und stellt dort nach PRINGSHEIM[7]) die Funktion $\log P(\theta(\varphi)) - \log R$ dar, so daß also gilt

$$h(z) = \sum_{n=1}^{\infty} (a_n + i b_n) z^n = \sum_{n=1}^{\infty} c_n z^n = \sum_{n=1}^{\infty} \frac{C_n}{R^n} z^n, \; C_n = A_n + i B_n.$$

Es ist dann mit $w = \varrho e^{i\theta}$

$$\log \varrho = \log \varrho(r, \varphi) = \log R + \sum_{n=1}^{\infty} \left( A_n \left(\frac{r}{R}\right)^n \cos n\varphi + B_n \left(\frac{r}{R}\right)^n \sin n\varphi \right)$$

$$\theta = \theta(r, \varphi) = \varphi + \sum_{n=1}^{\infty} \left( -B_n \left(\frac{r}{R}\right)^n \cos n\varphi + A_n \left(\frac{r}{R}\right)^n \sin n\varphi \right).$$

Da die Hilfsabbildungen so gewählt werden, daß $\Gamma_w$ wenig von einem Kreise abweicht, braucht man von den Reihen nur wenige Glieder mitzunehmen, so daß man auf diese Weise relativ bequem die $w$-Bilder von $z$ berechnen kann. Die Fourierkoeffizienten $A_n$ und $B_n$ brauchen nicht mehr besonders berechnet zu werden, wenn man, was oft rechnerische Vorteile bietet, die Auswertung der Integrale $\frac{1}{2\pi}\int_0^{2\pi} f_n(\alpha) \operatorname{ctg} \frac{\alpha - \varphi}{2} d\alpha$ auf die Fourierentwicklung der Funktion $f_n(\alpha)$ stützt.

---

[7]) PRINGSHEIM, A.: Über das Verhalten von Potenzreihen auf dem Konvergenzkreise. Sitzgsber. bayer. Akad. Wiss. **30** (1900).

*(Eingegangen am 25. Mai 1949.)*

Math. Annalen, Bd. 122, S. 14–15 (1950).

# Eine Formel der formalen Operatorenrechnung *).

Von

Gustav Herglotz in Göttingen.

Sind $X, Y, Z \ldots$ Elemente eines (nichtkommutativen) Ringes mit Einselement $E$, zu dem alle mit ihnen gebildeten Potenzreihen gehören, so gilt:

Ist $YX - XY = Z$ vertauschbar mit $X$ und $Y$,

so ist

$$e^{tXY} = \sum_0^\infty {}_\nu \frac{1}{\nu!} X^\nu Y^\nu \left(\frac{e^{tZ}-1}{Z}\right)^\nu,$$

insbesondere also für $YX - XY = E$ (Vertauschungsregel der Quantentheorie) in diesem Sinne

$$e^{tXY} = \sum_0^\infty {}_\nu \frac{1}{\nu!} (e^t - 1)^\nu X^\nu Y^\nu$$

und daher

$$e^{tXY} = E \text{ für } t = 2n\pi i.$$

Zum Beweise stelle man erst fest, daß es feste Zahlen $c_{\nu\mu}$ ($c_{\nu\mu} = 0$ für $\nu > \mu$) gibt, so daß

$$\text{für } \mu \geqq 1 : (XY)^\mu = \sum_1^\infty {}_\nu c_{\nu\mu} X^\nu Y^\nu Z^{\mu-\nu}.$$

Dies ist für $\mu = 1$ mit $c_{11} = 1$ trivialerweise erfüllt. Weiter bemerke man, daß

$$YX^\nu = X^\nu Y + \nu X^{\nu-1} Z$$

ist.

Danach hat man

$$(XY)^{\mu+1} = XY \cdot (XY)^\mu = \sum_1^\infty {}_\nu c_{\nu\mu} X \cdot YX^\nu \cdot Y^\nu \cdot Z^{\mu-\nu}$$

$$= \sum_1^\infty {}_\nu c_{\nu\mu} X^{\nu+1} Y^{\nu+1} Z^{\mu-\nu} + \sum_1^\infty {}_\nu \nu c_{\nu\mu} X^\nu Y^\nu Z^{\mu-\nu+1}$$

und ersieht, daß die $c_{\nu\mu}$ nur rekursiv für $\mu = 1, 2, 3, \ldots$ so zu wählen sind, daß der letzte Ausdruck gleich ist

$$= \sum_1^\infty {}_\nu c_{\nu\mu+1} X^\nu Y^\nu Z^{\mu-\nu+1}$$

Dies ist sicher der Fall, wenn es bei Ersetzung der hier auftretenden $\mu + 1$ Elemente $X^\nu Y^\nu Z^{\mu-\nu+1}$ ($\nu = 1, 2, \ldots, \mu + 1$) durch $s^\nu$ identisch in $s$ statthat.

---

*) Diese Bemerkung war in einer Mappe von Manuskripten enthalten, die Herrn C. L. Siegel zu seinem 50. Geburtstag am 31. Dezember 1946 von seinen Göttinger Freunden und Kollegen überreicht wurde.

Setzt man nun

$$\text{für } \mu \geqq 1 \; : \; f_\mu(s) = \sum_1^\infty {}_\nu \, c_{\nu,\mu} s^\nu, \quad f_0(s) = 1\,,$$

so erhält man

$$\mu \geqq 0 \; : \; f_{\mu+1}(s) = s f_\mu(s) + s f'_\mu(s) = e^{-s} \frac{d}{d\sigma}\left(e^s f_\mu(s)\right), \; \sigma = \lg s$$

somit

$$\mu \geqq 0 \; : \; f_\mu(s) = e^{-s} \frac{d^\mu e^s}{d\sigma^\mu}$$

und

$$\sum_0^\infty {}_\mu \frac{t^\mu}{\mu!} f_\mu(s) = e^{(e^t-1)s}$$

also

$$\sum_1^\infty {}_\mu \frac{c_{\nu\mu}}{\mu!} t^\mu = \frac{1}{\nu!}(e^t - 1)^\nu, \quad \nu \geqq 1$$

Damit aber wird sofort

$$e^{tXY} = E + \sum_1^\infty {}_\mu \frac{t^\mu}{\mu!}(XY)^\mu = E + \sum_1^\infty {}_\mu \sum_1^\infty {}_\nu \frac{t^\mu}{\mu!} c_{\nu\mu} X^\nu Y^\nu Z^{\mu-\nu} =$$

$$= E + \sum_1^\infty {}_\nu \frac{1}{\nu!}\left(\frac{e^{tZ}-1}{Z}\right)^\nu X^\nu Y^\nu$$

*(Eingegangen am 14. Mai 1949.)*

Math. Annalen, Bd. 122, S. 16—34 (1950).

# Über Matrixfunktionen.

Von

H. Richter in Haltingen (Baden).

## § 1. Einleitung.

Es kommt in den physikalischen Anwendungen oft vor, daß die reell- oder komplexwertigen Komponenten einer quadratischen $n$-reihigen Matrix $B$ eindeutige stetige Funktionen sind der Komponenten einer anderen quadratischen Matrix $A$ des gleichen Grades, derartig, daß diese Abhängigkeit invariant ist gegen affine Änderung des Koordinatensystems. Ein Beispiel dieser Art wird durch die Abhängigkeit des Spannungstensors vom Deformationstensor in der allgemeinen Elastizitätstheorie isotroper Medien gegeben, wenn diese beiden Tensoren geeignet gewählt werden[1]). Andererseits weiß man, daß es möglich ist, ganze analytische Funktionen auf Matrizen anzuwenden, da man durch formales Einsetzen einer Matrix $A$ anstelle der unabhängigen Variablen stets eine konvergente Reihe erhält, und daß diese Art der Abhängigkeit ebenfalls gegen die Transformation mit einer beliebigen dritten Matrix invariant ist, so daß diese funktionelle Abhängigkeit einen Spezialfall der zuerst genannten Abhängigkeit bildet. Es erscheint daher zweckmäßig, zunächst allgemein nach der Gestalt der invarianten Abhängigkeit zweier Matrizen zu fragen und erst dann zu untersuchen, wann diese Abhängigkeit als Anwendung einer gewöhnlichen Funktion auf Matrizen aufgefaßt werden kann. Dementsprechend unterscheiden wir im folgenden die Begriffe der „allgemeinen Matrixfunktion" und der „gewöhnlichen Matrixfunktion". Es wird sich dabei zeigen, daß diese Unterscheidung nur dann wesentlich ist, wenn wir eine bestimmte funktionelle Abhängigkeit auf variabel gedachtes $A$ anwenden; nicht dagegen, wenn wir die Gesamtheit aller möglichen Funktionen eines festen $A$ betrachten. Von den gewöhnlichen Matrixfunktionen dürften das größte Interesse diejenigen beanspruchen, die durch analytische (jedoch nicht notwendig ganze) Funktionen vermittelt werden: „analytische Matrixfunktionen". Wir werden in § 2 zunächst den Begriff der allgemeinen eindeutigen Matrixfunktion klären und Normaldarstellungen dafür angeben. Für spezielle $A$ wird die Matrixfunktion durch einen Grenzprozeß definiert, was beim Zusammenfallen von Eigenwerten von $A$ zu Mehrdeutigkeiten Anlaß geben kann.

In § 3 wird die Anwendung gewöhnlicher und insbesondere analytischer Funktionen $f(x)$ auf Matrizen besprochen. Es zeigt sich, daß bei analytischem $f(x)$ der Wert von $f(A)$ auch für die $A$ mit mehrfachen Eigenwerten eindeutig bestimmt ist und bei Kenntnis der Eigenwerte von $A$ und ihrer Vielfachheiten in einfacher Weise berechnet werden kann. Eine notwendige und hinreichende Bedingung für die Realität von $f(A)$ bei reellem $A$ wird abgeleitet.

---

[1]) Richter, H.: Verzerrungstensor, Verzerrungsdeviator und Spannungstensor bei endlichen Formänderungen. Z. ang. Math. Mech. **29**, 65—75 (1949).

§ 4 behandelt die Lösung der Matrixgleichung $A = \varphi(B)$ bei bekanntem $A$ und eindeutigem analytischen $\varphi$. Die allgemeine Gestalt von $B$ wird angegeben und für den Fall $\varphi(x) = x^2$, also $B = \sqrt{A}$ demonstriert.

In § 5 wird schließlich für analytische Matrixfunktionen $f(A)$ der Zusammenhang des Differentials mit $f'(A)$ behandelt und ein Analogon zu der gewöhnlichen Beziehung $d\,f = f' \cdot d\,x + 0\,((d\,x)^2)$ abgeleitet.

*Bezeichnungen.* Zur Matrix $A$ mit den Komponenten $a_{ik} = (A)_{ik}$ ist $\bar{A}$ die an der Hauptdiagonale Gespiegelte, $|A|$ die Determinante, $\{A\}$ die Spur. Für die Eigenwerte $\lambda_\nu$ von $A$ sind die Potenzsummen: $s_m = \sum\limits_\nu \lambda_\nu^m = \{A^m\}$. $E$ ist die Einheitsmatrix. $(e_\nu)$ ist eine Diagonalmatrix mit den Elementen $e_\nu$. $n$ ist der Grad von $A$.

In Übereinstimmung mit der üblichen Konvention beim Einsetzen von Matrizen in Polynome setzen wir $A^0 = E$ für *jedes* $A$.

$a$ und $a^*$ sind konjugiert-komplexe Zahlen.

## § 2. Die allgemeine stetige Matrixfunktion.

Sind vermöge einer Rechenvorschrift die Komponenten der Matrix $B$ eindeutige stetige Funktionen der variabel gedachten Komponenten der Matrix $A$ in einem offenen Teilbereich mit teilweise hinzugenommenen Randpunkten des $n^2$-dimensionalen Raumes der Komponenten von $A$, dann bedeutet es keine wesentliche Einschränkung, die Randpunkte zunächst wegzulassen, da die Werte von $B$ auf denselben stetig aus inneren Stellen gewonnen werden können. Wir können aus dem gleichen Grunde die geringerdimensionale Menge aller $A$ streichen, die zusammenfallende Eigenwerte haben. Wir bilden daher die folgende

*Definition 1*: Es sei $\mathfrak{B}_0$ ein offener Bereich von Matrizen $A$, die lauter verschiedene Eigenwerte haben. Mit $A$ seien auch alle äquivalenten $CAC^{-1}$ zu $A$ in $\mathfrak{B}_0$ enthalten. Zu jedem $A \in \mathfrak{B}_0$ sei stetig ein $B$ definiert. Dann heißt $B$ eine allgemeine stetige Matrixfunktion von $A$ in $\mathfrak{B}_0$, symbolisch

$$B = F\,u\,(A), \tag{2.1}$$

wenn für jedes $C$ mit $|C| \neq 0$ gilt:

$$CBC^{-1} = F\,u\,(CAC^{-1}). \tag{2.2}$$

Man sieht, daß der Begriff der allgemeinen Matrixfunktion eine Verallgemeinerung des Begriffes der Invariante darstellt.

Um nun die oben ausgeschlossenen $A$ mitzuerfassen, ergänzen wir durch

*Definition 2*: Liegt $A$ in der abgeschlossenen Hülle von $\mathfrak{B}_0$, dann gehört $A$ mit dem Funktionswert $F\,u\,(A) = B$ zum Definitionsbereich $\mathfrak{B}$ der Matrixfunktion, wenn es eine stetige Folge[2]) $A(t)$, $t > 0$, gibt mit:

$$A(t) \in \mathfrak{B}_0; \qquad \lim_{t\to 0} A(t) = A; \qquad \lim_{t\to 0} F\,u\,(A(t)) = B.$$

Hierbei kann es vorkommen, daß $F\,u\,(A)$ abhängig wird von der definierenden Folge $A(t)$. Wir unterscheiden daher gemäß

[2]) Hierbei kann an eine stetige Abhängigkeit der $A$ von dem reellen Parameter $t$ oder auch an eine konvergente Folge $A(t_n)$ mit $t_n \to 0$ gedacht werden.

*Definition 3*: Ein $A \in \mathfrak{B}$ ist eine *reguläre* Stelle der Matrixfunktion, wenn in Definition 2 $B$ unabhängig ist von der Folge $A(t)$.

Anderenfalls ist $A$ eine *richtungssinguläre* Stelle.

Da $F u(A)$ in $\mathfrak{B}_0$ stetig erklärt war, ist automatisch jedes $A \in \mathfrak{B}_0$ eine reguläre Stelle.

*Beispiel 1*: Für $F u(A) = \dfrac{A \cdot |A|}{|A|}$ besteht $\mathfrak{B}_0$ aus allen Matrizen mit lauter verschiedenen Eigenwerten und $|A| \neq 0$. $\mathfrak{B}$ enthält alle Matrizen. Jedes $A$ ist eine reguläre Stelle mit $F u(A) = A$.

*Beispiel 2*: $F u(A) = \sqrt[n]{|A|^{-1}} \cdot A$ hat das gleiche $\mathfrak{B}_0$ wie in Beispiel 1. Ist ein $A$ beliebig mit $|A| \neq 0$, dann erhalten wir aus Definition 2 unabhängig von $A(t)$ stets $B = \sqrt[n]{|A|^{-1}} \cdot A$. Ist $A \neq 0$ mit $|A| = 0$, so existiert $\lim\limits_{t \to 0} F u(A(t))$ für keine Folge $A(t)$. Ist schließlich $A = 0$, so kann aus Definition 2 für $B$ eine beliebige Matrix mit der Determinante 1 erhalten werden. $\mathfrak{B}$ besteht dann aus allen Matrizen mit $|A| \neq 0$ und der Matrix $A = 0$, wobei die letztere die einzige richtungssinguläre Stelle ist.

Wir können nun Definition 2 abrunden durch

*Satz 1*: Sei in $\mathfrak{B}$ enthalten eine stetige Folge $A(t)$ mit je einem $F u(A(t))$, $\lim\limits_{t \to 0} A(t) = A$ und $\lim\limits_{t \to 0} F u(A(t)) = B$. Dann ist $A \in \mathfrak{B}$ mit $F u(A) = B$.

*Beweis*: Nach Definition 2 gibt es $A(t, u) \in \mathfrak{B}_0$ mit $\lim\limits_{u \to 0} A(t, u) = A(t)$ und $\lim\limits_{u \to 0} F u(A(t, u)) = F u(A(t))$, $F u(A(t, u))$ gemäß Definition 1. Wir wählen zu jedem $t$ ein $u(t) > 0$ mit $u(t) \leq t$ und $\max |(A(t, u) - A(t))_{ik}| \leq t$ nebst $\max |(F u(A(t, u)) - F u(A(t)))_{ik}| \leq t$ für $u \leq u(t)$. $A(t) = A(t, u(t))$ erfüllt dann Definition 2.

Ebenso unmittelbar folgt aus unseren Definitionen, daß bei $A \in \mathfrak{B}$ mit $B = F u(A)$ auch jede Transformierte $A_1 = CAC^{-1}$ in $\mathfrak{B}$ liegt mit $F u(A_1) = CBC^{-1}$. Dabei sind die Transformierten zu regulären (richtungssingulären) Stellen wieder reguläre (richtungssinguläre) Stellen in $\mathfrak{B}$. Hieraus ergibt sich weiter, daß wir für die Untersuchung von $F u(A)$ stets $A$ in der Normalform

$$A' = \begin{pmatrix} A_1 & & 0 \\ & A_2 & \\ & & \ddots \\ 0 & & & A_m \end{pmatrix} \quad \text{mit } A_k = \begin{pmatrix} \lambda_k & 1 & & 0 \\ & \lambda_k & 1 & \\ & & \ddots & 1 \\ 0 & & & \lambda_k \end{pmatrix} \tag{2.3}$$

annehmen können, in die bekanntlich jedes $A$ affin transformiert werden kann.

Ist speziell $C$ mit $A$ vertauschbar und $|C| \neq 0$, so ist mit $B = F u(A)$ auch $CBC^{-1} = F u(A)$.

Für reguläre Stellen $A$ gibt es nach Definition 3 nur ein $F u(A)$. Also gilt bei $|C| \neq 0$:

*Satz 2a*: Für *reguläre* Stellen $A$ ist $F u(A)$ mit allen $C$ vertauschbar, die mit $A$ vertauschbar sind.

Bei $|C| = 0$ ist für kleine $u > 0$ sicher $|C + u E| \neq 0$ und $C + u E$ mit $A$ vertauschbar. Aus $(C + u E) F u(A) = F u(A) \cdot (C + u E)$ folgt bei $u \to 0$ die allgemeine Gültigkeit von Satz 2a. (Beispiel 2 zeigt, daß die

entsprechende Behauptung für richtungssinguläre Stellen falsch sein kann.) Da alle $A \in \mathfrak{B}_0$ regulär sind, ist also insbesondere für die in Definition 2 verwendete Folge $A(t)$: $A(t) \cdot Fu(A(t)) = Fu(A(t)) \cdot A(t)$. Der Grenzübergang zu $t = 0$ liefert

*Satz 2b*: Für *alle* $A \in \mathfrak{B}$ ist $Fu(A)$ mit $A$ vertauschbar.

Nach diesen einleitenden Sätzen wollen wir die Art der durch eine Matrixfunktion geschaffenen Abhängigkeit klären. Hierzu gehen wir von den $A \in \mathfrak{B}_0$ aus. Es gibt für jedes solches $A$ wenigstens ein $C$, das $A$ in Diagonalform $CAC^{-1} = (\lambda_\nu)$ transformiert. Nach (2.2) und Satz 2b ist dann $CBC^{-1} = (\mu_\nu)$ mit

$$\mu_\nu = g_\nu(\lambda_\nu; \lambda_1, \ldots, \lambda_{\nu-1}, \lambda_{\nu+1}, \ldots, \lambda_n),$$

wo $g_\nu$ in den Variablen $\lambda_1, \ldots, \lambda_{\nu-1}, \lambda_{\nu+1}, \ldots, \lambda_n$ symmetrisch sein muß, da man durch geeignete Transformation die entsprechenden Einheitsvektoren vertauschen kann, ohne gemäß (2.2) den Wert von $\mu_\nu$ zu verändern. $g_\nu$ hängt daher von $\lambda_\nu$ und den ersten $(n-1)$ Potenzsummen der $\lambda_\mu \neq \lambda_\nu$ ab. Da man dieselben aber auch als Funktionen von $\lambda_\nu$ und den entsprechenden Potenzsummen aller $\lambda_\nu$ schreiben kann, ist somit

$$\mu_\nu = h(\lambda_\nu; s_1, s_1, \ldots, s_{n-1}).$$

Hierbei ist der Index $\nu$ bei $h$ weggelassen worden, da wegen der Invarianz gegen Vertauschung aller Grundvektoren die Funktion $h$ für alle $\nu$ die gleiche sein muß. $h$ ist eine gewöhnliche stetige Funktion, die durch $Fu$ wegen der Unabhängigkeit der Variablen $\lambda_\nu, s_1, \ldots, s_{n-1}$ eindeutig bestimmt ist und die Matrixfunktion vermittelt. Ist umgekehrt eine stetige Funktion $h(\lambda; s_1, \ldots, s_{n-1})$ vorgegeben, so können wir für alle Matrizen, deren Eigenwerte $\lambda_\nu$ verschieden sind und für welche die Argumente $\lambda_\nu, s_1, \ldots, s_{n-1}$ im Inneren des Definitionsbereiches von $h$ liegen, eine Matrix $B$ durch $B = C^{-1} \times (h(\lambda_\nu; s_1, \ldots, s_{n-1})) \cdot C$ bilden. $B$ ist dann eindeutig bestimmt und hängt stetig von $A$ ab, da bei verschiedenen $\lambda_\nu$ ein stetig verändertes $A$ auch stetige Veränderungen der $\lambda_\nu$ und von $C$ nach sich zieht. Damit sind die Bedingungen von Definition 1 erfüllt.

Es ist daher zweckmäßig, anstelle des allgemeinen Symbols $Fu$ die spezielle Gestalt der Abhängigkeit durch die folgende Schreibweise zum Ausdruck zu bringen:

(2.4) $$B = h(A; s_1, \ldots, s_{n-1}) \quad \textit{1. Normaldarstellung.}$$

Durch (2.4) soll gleichzeitig das soeben angegebene Verfahren der Berechnung von $B$ symbolisiert werden. Weiter setzt die 1. Normaldarstellung die Eigenwerte von $B$ in Evidenz.

Haben wir eine Matrix $A$ mit teilweise zusammenfallenden Eigenwerten, die sich jedoch in Diagonalform $CAC^{-1} = (\lambda_\nu)$ bringen läßt („symmetrisierbare Matrix") und liegen die Argumente $\lambda_\nu, s_1, \ldots, s_{n-1}$ im stetigen Definitionsbereich von $h$, dann können wir jedenfalls wie oben $B$ bilden. $A$ liegt dann in der abgeschlossenen Hülle von $\mathfrak{B}_0$ und $B$ erscheint als Limes von $Fu(A(t))$ mit $A(t) \in \mathfrak{B}_0$. Solche $A$ liegen also jedenfalls in $\mathfrak{B}$. Wir wollen uns nun überzeugen, daß jedoch $A$ durchaus eine richtungssinguläre Stelle sein kann, so daß (2.4) nicht den gesamten Wertevorrat von $Fu(A)$ für die $A \in \mathfrak{B}$ mit zusammenfallenden Eigenwerten liefert. Dieses Verhalten zeigt

*Beispiel 3*: Es sei

$$f(1 + r e^{i\varphi}) = \begin{cases} 1 + \sqrt{r}\, e^{i\frac{\varphi}{2}} & \text{für } |\varphi| \le \frac{\pi}{2} \\ 1 + \sqrt{r}\, e^{\frac{i}{2}(3|\varphi| - \pi)\cdot \operatorname{sign}\varphi} & \text{für } \frac{\pi}{2} \le |\varphi| \le \pi. \end{cases}$$

$f(x)$ ist eine für alle $x$ stetige Funktion. Wir setzen $h(\lambda_\nu; s_1, \ldots, s_{n-1}) \equiv$ $\equiv f(\lambda_\nu)$. Durch direkte Anwendung von (2.4) finden wir bei $A = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} = E$:

$$F u(E) = E.$$

Benutzen wir aber die bei $t \to 0$ gegen $E$ konvergente Folge

$$A(t) = \begin{pmatrix} 1 + it^2 & 0 \\ t\sqrt{2} & 1 - it^2 \end{pmatrix} = V \cdot \begin{pmatrix} 1 + it^2 & 0 \\ 0 & 1 - it^2 \end{pmatrix} \cdot V^{-1} \text{ mit } V = \begin{pmatrix} it\sqrt{2} & 0 \\ 1 & 1 \end{pmatrix},$$

so ist

$$F u(A(t)) = V \cdot \begin{pmatrix} 1 + te^{i\frac{\pi}{4}} & 0 \\ 0 & 1 + te^{-i\frac{\pi}{4}} \end{pmatrix} V^{-1} = \begin{pmatrix} 1 + te^{i\frac{\pi}{4}} & 0 \\ 1 & 1 + te^{-i\frac{\pi}{4}} \end{pmatrix}$$

und damit bei $t \to 0$:

$$F u(E) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Liegt mit $A$ auch $\overline{A}$ in $\mathfrak{B}_0$, so ist bei $A = C^{-1}(\lambda_\nu)\, C$ und $B = C^{-1}(\mu_\nu)\, C$ wegen (2.2):

$$F u(\overline{A}) = F u(\overline{C}(\lambda_\nu)\overline{C}^{-1}) = \overline{C}\, F u(\lambda_\nu)\, \overline{C}^{-1} = \overline{C}(\mu_\nu)\overline{C}^{-1} = \overline{C^{-1}(\mu_\nu)\, C} = \overline{B}$$

Wir können daher gegebenenfalls den Bereich $\mathfrak{B}_0$ dadurch erweitern, daß wir alle gespiegelten Matrizen hinzunehmen und dabei $F u(\overline{A}) = \overline{F u(A)}$ setzen. Anschließend können wir dann unter Erhaltung dieser Beziehung gemäß Definition 2 den Definitionsbereich $\mathfrak{B}$ bilden, der dann ebenfalls mit jeder Matrix auch ihre Gespiegelte enthält. Es gilt daher

*Satz 3*: Zu $\mathfrak{B}$ dürfen die Gespiegelten aller Matrizen $A \in \mathfrak{B}$ hinzugenommen werden mit $F u(\overline{A}) = \overline{F u(A)}$.

Liegt $A$ in $\mathfrak{B}_0$ und ist $A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$, so ist gemäß der durch die erste Normaldarstellung gegebenen Rechenvorschrift $F u(A) = B = \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}$, wobei wir $B_1$ als eine Matrixfunktion von $A_1$ auffassen können, die außerdem noch die Invarianten von $A_2$ als Parameter enthält. Für richtungssinguläre Stellen $A$ zeigt dagegen Beispiel 3, daß aus der Reduziertheit von $A$ nicht die von $B$ folgt. Richtig ist diese Behauptung jedoch, wenn $A_1$ und $A_2$ eigenwertfremd sind, da dann die Reduziertheit von $B$ sich aus Satz 2b ergibt. Um nun zu sehen, daß auch in diesem Falle $B_1$ aus einer Matrixfunktion von $A_1$ mit den Invarianten von $A_2$ als Parametern gewonnen wird, beweisen wir zunächst

*Hilfssatz 1*: Ist gegeben eine Folge $A(t)$ bei $\lim\limits_{t\to 0} A(t) = A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$ mit eigenwertfremden $A_1$ und $A_2$, dann gibt es eine Folge $T(t)$ mit

$$A(t) = T(t) \cdot \begin{pmatrix} A_1(t) & 0 \\ 0 & A_2(t) \end{pmatrix} T^{-1}(t); \ \lim_{t \to 0} T(t) = E; \ \lim_{t \to 0} A_\nu(t) = A_\nu.$$

*Beweis:* Der Grad von $A_\nu$ sei $m_\nu$: $m_1 + m_2 = n$. Ist nun $\varphi_\nu(y) = 0$ die charakteristische Gleichung von $A_\nu$, so entspricht $\varphi_\nu(y)$ wegen der Eigenwertfremdheit der $A_\nu$ für kleine $t$ eindeutig einem Faktor $\varphi_\nu(y, t)$ der charakteristischen Gleichung von $A(t)$. Bilden wir $C_\nu(t) = \varphi_\nu(A(t), t)$, so hat $C_\nu(t)$ genau den Rang $n - m_\nu$. $C_\nu(t)$ bestimmt also vermöge $C_\nu(t) \cdot \mathfrak{x} = 0$ einen $m_\nu$-dimensionalen Unterraum $U_\nu(t)$ als Eigenraum von $A(t)$, der wegen der Konstanz des Ranges bei $t \to 0$ in den durch $\varphi_\nu(A) \cdot \mathfrak{x} = 0$ definierten Unterraum $U_\nu$ übergeht, der durch die ersten $m_1$, bzw. letzten $m_2$, Grundvektoren $\mathfrak{e}_i$ aufgespannt wird. Es gibt also in $U_\nu(t)$ $m_\nu$ linear unabhängige Vektoren $\mathfrak{f}_i(t)$ mit $\lim_{t \to 0} \mathfrak{f}_i(t) = \mathfrak{e}_i$. Alle $\mathfrak{f}_i(t)$ zusammen spannen den ganzen Raum auf. Das gesuchte $T(t)$ ist die Matrix, die $\mathfrak{e}_i$ in $\mathfrak{f}_i(t)$ transformiert.

Wir haben nun

*Satz 4*: Ist $A \in \mathfrak{B}$ und $A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$ mit eigenwertfremden $A_\nu$, so ist $Fu(A) = \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}$, wo $B_1$ Matrixfunktion von $A_1$ mit den Invarianten von $A_2$ als Parametern ist: $B_1 = Fu(A_1; \{A_2^k\})$.

*Beweis*: Für $A \in \mathfrak{B}_0$ wurde der Beweis oben geführt. Für die übrigen $A \in \mathfrak{B}$ wird $Fu(A)$ gemäß Definition 2 mit Hilfe einer Folge $A(t)$ gebildet, auf die Hilfssatz 1 anwendbar ist. Es ist dann wegen (2.2):

$$Fu(A(t)) = T(t) \cdot \begin{pmatrix} Fu(A_1(t); \{A_2^k(t)\}) & 0 \\ 0 & Fu(A_2(t); \{A_1^k(t)\}) \end{pmatrix} T^{-1}(t),$$

woraus bei $t \to 0$ unmittelbar die Behauptung folgt.

In (2.4) kommt die letzte unabhängige Invariante $s_n$ von $A$ nicht mit vor. Man kann sie aber mitnehmen und allgemeiner $B = g(A; s_1, \ldots, s_n)$ schreiben. Doch ist diese Darstellung dann nicht mehr eindeutig. Da man umgekehrt $s_n$ als Funktion eines $\lambda_\nu$ und der $s_1, \ldots, s_{n-1}$ schreiben kann, läßt sich stets die eindeutige Normaldarstellung (2.4) zurückgewinnen. Man kann nun diesen Sachverhalt ausnutzen, um durch die Einführung von $s_n$ die Abhängigkeit der Funktion $g$ von ihrer ersten Variablen möglichst einfach zu gestalten. Hierzu bestimmen wir für $A \in \mathfrak{B}_0$ die Größen $c_0, c_1, \ldots, c_{n-1}$ aus dem Gleichungssystem:

$$\text{(2.5)} \qquad \sum_{k=0}^{n-1} c_k \cdot \lambda_\nu^k = h(\lambda_\nu; s_1, \ldots, s_{n-1}) \qquad \nu = 1, 2, \ldots, n.$$

Die $c_k$ sind offensichtlich invariant gegen Permutationen der $\lambda_\nu$ und können somit als Funktionen von $s_1, \ldots, s_n$ aufgefaßt werden:

$$c_k = c_k(s_1, \ldots, s_n).$$

Es wird dann nach der durch (2.4) symbolisierten Rechenvorschrift:

$$B = C^{-1} \cdot \left( \sum_{k=0}^{n-1} c_k \lambda_\nu^k \right) C = \sum_{k=0}^{n-1} c_k \cdot C^{-1}(\lambda_\nu^k) C = \sum_{k=0}^{n-1} c_k \cdot [C^{-1}(\lambda_\nu) C]^k$$

oder

$$\text{(2.6)} \qquad B = \sum_{k=0}^{n-1} c_k(s_1, \ldots, s_n) \cdot A^k. \quad \textit{2. Normaldarstellung.}$$

$Fu(A)$ kann also stets als ganz-rationale Funktion von $A$ vom Grade $n-1$ aufgefaßt werden, wobei die Koeffizienten Invarianten von $A$ sind. Umgekehrt liefert (2.6) mit beliebigen stetigen Invarianten $c_k$ stets eine stetige Matrixfunktion für einen geeigneten Bereich $\mathfrak{B}_0$.

Die 2. Normaldarstellung gilt zunächst wieder nur für die $A \in \mathfrak{B}_0$. Für andere $A$ können bei Approximation mit Hilfe einer Folge $A(t)$ die gemäß (2.5) gebildeten $c_k(t)$ konvergieren. Dann gehört zwar $A$ zu $\mathfrak{B}$, kann aber durchaus eine richtungssinguläre Stelle sein, so daß (2.6) dann nicht alle Werte von $Fu(A)$ wiedergibt. Dies folgt auch daraus, daß ein gemäß (2.6) gebildetes $B$ automatisch mit allen $C$ vertauschbar ist, die mit $A$ vertauschbar sind.

Da jede Matrix ihrer charakteristischen Gleichung genügt, kann für jedes $m$ die Funktion $A^m$ durch $A^0, A^1, \ldots, A^{n-1}$ und $s_1, \ldots, s_n$ dargestellt werden. (2.6) ist die Verallgemeinerung dieses Satzes für beliebige Matrixfunktionen.

Wir hatten gesehen, daß wir die erste Normaldarstellung (2.4) auf alle $A$ aus $\mathfrak{B}$ anwenden dürfen, die symmetrisierbar sind und damit jedenfalls wenigstens einen Wert für $Fu(A)$ erhalten. Doch ist die Anwendung von (2.4) wegen der Transformation auf Hauptachsen sehr umständlich. Für alle symmetrisierbaren $A$ läßt sich nun aber (2.4) wie folgt umformen:

Es seien $\lambda_1, \ldots, \lambda_m$ mit $m \leq n$ die verschiedenen Eigenwerte von $A$; dann bilden wir die $m$ Hilfsmatrizen

$$(2.7) \qquad G_k = \prod_{i \neq k} \frac{A - \lambda_i E}{\lambda_k - \lambda_i}, \qquad k = 1, \ldots, m.$$

Ist nun $CAC^{-1} = (\lambda_\nu)$, so wird $CG_kC^{-1}$ eine Diagonalmatrix, die nur an den Stellen, wo in $(\lambda_\nu)$ der Wert $\lambda_k$ stand, eine 1 besitzt, sonst aber lauter Nullen hat. Es ist daher gemäß (2.4):

$$C\,B\,C^{-1} = \sum_{k=1}^{m} \mu_k \cdot C\,G_k\,C^{-1}$$

und damit

$$(2.8) \qquad \begin{cases} h(A; s_1, \ldots, s_{n-1}) = \sum_{k=1}^{m} \mu_k \cdot G_k \text{ mit} \\ \mu_k = h(\lambda_k; s_1, \ldots, s_{n-1}) \text{ und} \\ G_k = \prod_{i \neq k} \dfrac{A - \lambda_i E}{\lambda_k - \lambda_i}. \end{cases} \qquad \textit{3. Normaldarstellung}$$

Die 3. Normaldarstellung ist gültig für alle $A \in \mathfrak{B}_0$ und für alle übrigen symmetrisierbaren $A \in \mathfrak{B}$, für die sie aber evtl. nicht alle Werte liefert. Für nichtsymmetrisierbare $A$ liefert sie falsche Resultate. Sie ist besonders bequem für numerische Rechnungen, vor allem dann, wenn es sich darum handelt, von einer festen Matrix $A$ verschiedene Funktionen zu berechnen, da in diesem Falle die Aufstellung der $G_k$ nur einmal zu geschehen braucht.

## § 3. Gewöhnliche, insbesondere analytische Matrixfunktionen.

Ist die Abhängigkeit $Fu(A)$ durch Anwendung der Potenzreihe einer ganzen analytischen Funktion $f(x)$ auf $A$ gegeben, so ist $\mu_\nu = f(\lambda_\nu)$, so daß in der eindeutig bestimmten ersten Normaldarstellung die explizite

Abhängigkeit von $s_1, \ldots, s_{n-1}$ entfällt. Es ist daher sinnvoll, ganz allgemein die Anwendung einer eindeutigen Funktion $f(x)$ auf Matrizen im Sinne der Vorschrift der ersten Normaldarstellung aufzufassen und dann einfach

$$B = f(A) \tag{3.1}$$

zu schreiben. Diese so gewonnene „gewöhnliche Matrixfunktion" heißt „analytische Matrixfunktion", wenn $f(x)$ an den Stellen der Eigenwerte der $A$ aus $\mathfrak{B}$ analytisch ist.

Betrachten wir eine Menge von allgemeinen Funktionen für ein und dieselbe Matrix $A$, so sind die $s_k$ in (2.4) feste Zahlen, so daß der Unterschied zwischen allgemeiner und gewöhnlicher Matrixfunktion verwischt wird. Die Unterscheidung ist jedoch wesentlich, sobald wir *eine* Matrixfunktion für variabel gedachtes $A$ untersuchen. Im Falle der analytischen Matrixfunktion sind aus der RIEMANNschen Fläche von $f(x)$ geeignete Stücke auszuschneiden, über denen dann $f(x)$ als eindeutige Funktion betrachtet werden kann, wodurch $\mathfrak{B}$ festgelegt wird. Solange $A$ verschiedene Eigenwerte hat, bedeutet die hierdurch geschaffene Einengung des Variabilitätsbereiches von $A$ keine wesentliche Einschränkung, da man ja die Schnitte geeignet legen kann, um alle Zweige von $f(x)$ benutzen zu können. Schwierigkeiten tauchen nur bei denjenigen $A$ auf, bei denen Eigenwerte zusammenfallen, da dann der durch Definition 2 geforderte Grenzübergang dazu führen kann, daß gleichen $\lambda_\nu$ verschiedene $\mu_\nu$ entsprechen. (Weiteres hierzu siehe in § 4.) Bei gewöhnlichen Matrixfunktionen ist in Satz 4 für die $A \in \mathfrak{B}_0$ einfach $F u(A_1; \{A_2^k\}) = f(A_1)$. Satz 4 gestattet daher jetzt die einfachere Formulierung von

*Satz 5*: Ist $A \in \mathfrak{B}$ und $A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$ mit eigenwertfremden $A_\nu$, so ist $f(A) = = \begin{pmatrix} f(A_1) & 0 \\ 0 & f(A_2) \end{pmatrix}$.

Damit ist die Bestimmung von $f(A)$ auf die $A$ mit lauter gleichen Eigenwerten $\lambda$ zurückgeführt. Wie Beispiel 3 zeigte, ist Satz 5 ohne die Einschränkung der Eigenwertfremdheit der $A_\nu$ falsch. Das in Beispiel 3 gezeigte Verhalten ist nun wesentlich dadurch bestimmt, daß $f(x)$ nicht analytisch ist. Wir führen daher jetzt die weitere Untersuchung von $f(A)$ bei $A$ mit lauter gleichen Eigenwerten $\lambda$ unter der Voraussetzung des analytischen Charakters von $f(x)$ an der Stelle $x = \lambda$ durch. Hierbei ist für die Beurteilung des analytischen Charakters von $f(x)$ auf die Schnitte Rücksicht zu nehmen, durch die $f(x)$ eindeutig gemacht wurde. Hat man z. B. die RIEMANNsche Fläche von $f(x) = \sqrt{x}$ längs der negativ-reellen Achse aufgeschnitten, so zählt diese einschließlich $x = 0$ nicht zu den Argumenten, wo $f(x)$ analytisch ist. Man kann aber durch geschicktes Legen der Schnitte erreichen, daß jedes $A$ erfaßt wird, für welches $f(x)$ im gewöhnlichen Sinne an den Stellen der Eigenwerte von $A$ regulär ist.
Schreiben wir nun:

$$f(x) = g(x - \lambda) + f(\lambda) \quad \text{mit } g(0) = 0,$$

so ist bei $t > 0$ wegen der Verschiedenheit der Eigenwerte $\lambda_\nu(t)$ der Matrix $A(t)$ aus der gegen $A$ konvergenten Folge:

$$f(A(t)) = g(A(t) - \lambda E) + f(\lambda) \cdot E. \tag{3.2}$$

$g(x)$ habe nun die für kleine $x$ konvergente Entwicklung:

$$g(x) = \sum_{k=1}^{\infty} d_k x^k = \sum_{k=1}^{N-1} d_k x^k + h(x) \quad \text{mit} \quad d_k = \frac{1}{k!} f^{(k)}(\lambda),$$

so daß wir erhalten

$$g(A(t) - \lambda E) = \sum_{k=1}^{N-1} d_k \cdot (A(t) - \lambda E)^k + h(A(t) - \lambda E). \tag{3.3}$$

Gemäß (2.6) ist:

$$h(A(t) - \lambda E) = \sum_{i=1}^{n-1} c_i(t) \cdot (A(t) - \lambda E)^i,$$

wo die $c_i(t)$ bestimmt sind durch (2.5):

$$\sum_{i=0}^{n=1} c_i(t)(\lambda_\nu(t) - \lambda)^i = h(\lambda_\nu(t) - \lambda).$$

Bei genügend großem $N$ folgt hieraus wegen $h(x) = O(x^N)$:

$$\lim_{t\to 0} c_i(t) = 0$$

und damit

$$\lim_{t\to 0} h(A(t) - \lambda E) = 0.$$

Aus (3.2) und (3.3) erhalten wir nün bei $t \to 0$ unabhängig von der definierenden Folge $A(t)$ den Wert für $f(A)$, so daß wir wegen $(A - \lambda E)^k = 0$ für $k \geq n$ haben

*Satz 6*: Ist $f(x)$ analytisch an der Stelle $x = \lambda$ des $n$-fachen Eigenwertes von $A$ (des Grades $n$), so ist $A \in \mathfrak{B}$; $A$ ist eine reguläre Stelle der Matrixfunktion mit

$$f(A) = \sum_{k=0}^{n-1} \frac{1}{k!} f^{(k)}(\lambda) \cdot (A - \lambda E)^k.$$

Wenn $A$ reduziert ist, so ist auch $f(A)$ reduziert, so daß also Satz 5 für analytische Matrixfunktionen auch ohne die Einschränkung der Eigenwertfremdheit richtig ist, wenn $f(x)$ an den Stellen der Eigenwerte von $A$ analytisch ist. Es sei nun $A'$ eine Matrix in der reduzierten Gestalt

$$A' = \begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_m \end{pmatrix},$$

wo die eigenwertfremden $A_\nu$ vom Grade $n_\nu$ mit den $n_\nu$-fachen Eigenwerten $\lambda_\nu$ sind. $f(x)$ sei analytisch für $x = \lambda_\nu$. Für jedes $\nu$ sind dann die Polynome $\prod_{\mu \neq \nu} (x - \lambda_\mu)^{n_\mu}$ und $(x - \lambda_\nu)^{n_\nu}$ teilerfremd, so daß wir ein Polynom $H_\nu(x)$ bestimmen können durch

$$H_\nu(x) \cdot \prod_{\mu \neq \nu} (x - \lambda_\mu)^{n_\mu} \equiv 1 \pmod{(x - \lambda_\nu)^{n_\nu}}. \tag{3.4}$$

Wir führen nun ein

$$G'_\nu = H_\nu(A') \cdot \prod_{\mu \neq \nu} (A' - \lambda_\mu E)^{n_\mu},$$

dann ist

$$(3.5) \qquad G'_\nu = \begin{pmatrix} 0 & & & & \\ & \ddots & & 0 & \\ & & 0 \; E_\nu \; 0 & & \\ & 0 & & \ddots & \\ & & & & 0 \end{pmatrix} \quad \text{und} \quad G_\nu A' = \begin{pmatrix} 0 & & & & \\ & \ddots & & 0 & \\ & & 0 \; A_\nu \; 0 & & \\ & 0 & & \ddots & \\ & & & & 0 \end{pmatrix}.$$

Nach den Sätzen 5 und 6 haben wir nun

$$f(A') = \sum_{\nu=1}^{m} G'_\nu \cdot \sum_{k=0}^{n_\nu - 1} \frac{1}{k!} f^{(k)}(\lambda_\nu) \cdot (A' - \lambda_\nu E)^k.$$

Durch Transformation mit einer beliebigen Matrix $C$ erhalten wir hieraus den

*Satz 7*: Hat $A$ die Eigenwerte $\lambda_\nu$ mit den Vielfachheiten $n_\nu$, und ist $f(x)$ an den Stellen $\lambda_\nu$ analytisch, so ist $A \in \mathfrak{B}$ mit dem eindeutig bestimmten Werte[3])

$$f(A) = \sum_{\nu=1}^{m} G_\nu \cdot \sum_{k=0}^{n_\nu - 1} \frac{1}{k!} f^{(k)}(\lambda_\nu) \cdot (A - \lambda_\nu E)^k,$$

wo

$$G_\nu = H_\nu(A) \cdot \prod_{\mu \neq \nu} (A - \lambda_\mu E)^{n_\mu} \text{ ist mit } H_\nu(x) \text{ gemäß (3.4)}.$$

*Bemerkung 7a*: Anstelle von $n_\nu$ können die Exponenten $m_\nu \leq n_\nu$ benutzt werden, für die gerade $\prod_\nu (A - \lambda_\nu E)^{m_\nu} = 0$ ist; bei symmetrisierbaren $A$ also $m_\nu = 1$. Im letzteren Falle ist dann

$$H_\nu(x) = \frac{1}{\prod_{\mu \neq \nu} (\lambda_\nu - \lambda_\mu)},$$

so daß Satz 7 ein Spezialfall von (2.8) wird[4]).

*Bemerkung 7b*: Der in Satz 7 angegebene Ausdruck für $f(A)$ kann noch modulo der charakteristischen Gleichung von $A$ reduziert werden, so daß man die Darstellung (2.6) erhält. In der Tat bleiben ja die $c_k$ als Lösung von (2.5) analytische Funktionen der Eigenwerte $\lambda_\nu$, auch wenn einige derselben zusammenrücken. Sie genügen dann dem aus (2.5) durch Grenzübergang entstehenden Gleichungssystem:

$$(3.6) \qquad \sum_{k=r}^{n-1} c_k \binom{k}{r} \lambda_\nu^{k-r} = \frac{1}{r!} f^{(r)}(\lambda_\nu)$$

---

[3]) Vgl. Fantappiè, L.: Le calcul des matrices, C. r. **186**, 619—621 (1928), wo durch Anwendung der Theorie der linearen analytischen Funktionale auf die Komponenten von $f(A)$ gezeigt wird, daß allein von $A$ abhängige Matrizen $H_{\nu k}$ existieren, so daß

$$f(A) = \sum_{\nu=1}^{m} \sum_{k=0}^{n_\nu - 1} H_{\nu k} f^{(k)}(\lambda_\nu)$$

wird.

[4]) Unter der Einschränkung, daß $f(x)$ eine ganze Funktion ist und keine Eigenwerte von $A$ zusammenfallen, bereits zu finden bei A. D. Michal, Matrix and Tensor Calculus, S. 18. New York 1947.

bei $r = 0, 1, \ldots, n_\nu - 1$ und $\nu = 1, 2, \ldots, m$; $n_\nu$ = Vielfachheit des Eigenwertes $\lambda_\nu$.

Aus Satz 7 können wir nun ein Kriterium dafür ableiten, daß bei reellem $A$ auch $f(A)$ reell ist. Von den Eigenwerten von $A$ seien in diesem Falle $\lambda_1, \ldots, \lambda_{2r}$ paarweise konjugiert komplex und $\lambda_{2r+1}, \ldots, \lambda_m$ reell. Es ist dann nach (3.4): $H_1 = H_2^*, \ldots, H_{2r-1} = H_{2r}^*$, während $H_{2r+1}$ bis $H_m$ reelle Polynome sind. Also ist wegen der Realität von $A$: $G_1 = G_2^*, \ldots, G_{2r-1} = G_{2r}^*$, und $G_{2r+1}$ bis $G_m$ sind reelle Matrizen. Notwendig und hinreichend dafür, daß $f(A)$ reell ist, ist also nach Satz 7, wenn wir gemäß Bemerkung 7a die $m_\nu$ anstelle der $n_\nu$ benutzen:

$$\sum_{\nu=1}^{m} \sum_{k=0}^{m_\nu - 1} \frac{1}{k!} (f^{(k)}(\lambda_\nu) - [f^{(k)}(\lambda_\nu^*)]^*) \cdot G_\nu \cdot (A - \lambda_\nu E)^k = 0\,.$$

An (3.5) sehen wir nun, daß die Matrizen $G_\nu \cdot (A - \lambda_\nu E)^k$ mit $0 \leq k \leq m_\nu - 1$ voneinander linear unabhängig sind. Damit haben wir den

*Satz 8*: Notwendig und hinreichend, daß bei reellen $A$ die analytische Matrixfunktion $f(A)$ ebenfalls reell ist, ist

$$f^{(k)}(\lambda_\nu) = [f^{(k)}(\lambda_\nu^*)]^* \text{ für } k = 0, 1, \ldots, m_\nu - 1,$$

wenn $\prod_\nu (x - \lambda_\nu)^{m_\nu}$ das Polynom niedrigsten Grades ist, welches $A$ annulliert.

*Bemerkung 8a*: Ist $n_\nu$ die Vielfachheit von $\lambda_\nu$, so ist $m_\nu \leq n_\nu$, so daß die Bedingung von Satz 8 für $k = 0, 1, \ldots, n_\nu - 1$ jedenfalls hinreichend ist. Bei symmetrisierbarem $A$ ist $m_\nu = 1$, so daß nur $f(\lambda_\nu) = (f(\lambda_\nu^*))^*$ zu fordern ist.

*Bemerkung 8b*: Ist $f(x)$ eine reelle analytische Funktion, d. h. ist $f(x)$ reell auf einem Stück der reellen Achse, und wird derjenige Zweig von $f(x)$ gewählt, für den $f(x^*) = (f(x))^*$ ist, so sagt Satz 8 aus, daß $f(A)$ genau dann reell ist, wenn $f(\lambda_\nu)$ reell ist für die reellen unter den $\lambda_\nu$. So ist z. B. $\ln A$ reell bei den reellen $A$, deren reelle Eigenwerte (falls vorhanden) positiv sind. Satz 8 zeigt aber, daß die Möglichkeit der Realität von $f(A)$ im allgemeinen größer ist, da die Bedingung von Satz 8 für bestimmte $\lambda_\nu$ ganz zufällig bis zu $k = m_\nu - 1$ erfüllt sein kann, während die höheren Ableitungen von $f(x)$ an dieser Stelle nicht mehr dieser Bedingung zu genügen brauchen. Dies hätte zur Folge, daß für jedes reelle $D$ dann $f(A + t\,D)$ auch für kleines $t > 0$ nicht mehr reell ist, so daß man $A$ in Übertragung einer bei den reellen Funktionen üblichen Sprechweise als „isolierte Matrix für $f(A)$" anzusprechen hat.

## § 4. Die Umkehrung analytischer Matrixfunktionen.

Wir hatten in § 3 für die Definition der analytischen Matrixfunktion eine Aufschneidung der RIEMANNschen Fläche von $f(x)$ vorgeschrieben, um zu einer eindeutigen Matrixfunktion zu gelangen. Man wird nun die Frage aufwerfen, welches die Gesamtheit aller $f(A)$ ist, wenn wir bei der gemäß Definition 2 zu erfolgenden Bestimmung von $f(A)$ mit Hilfe einer Folge $A(t)$ die volle Bewegungsfreiheit der Eigenwerte von $A(t)$ auf der RIEMANNschen Fläche zulassen. Solange die Eigenwerte von $A$ alle untereinander und von den singulären Stellen von $f(x)$ verschieden sind, kann sich natürlich nichts Neues ergeben, da wir durch geeignete Aufschneidung stets

erreichen können, daß $A \in \mathfrak{B}_0$ ist. Wenn dagegen $A$ teilweise zusammenfallende Eigenwerte hat und $\lim_{t \to 0} A(t) = A$ ist, wo $A(t)$ für $t > 0$ die verschiedenen Eigenwerte $\lambda_\nu(t)$ hat, so bilden die $\lambda_\nu(t)$ „Bahnkurven" in der $x$-Ebene mit teilweise zusammenfallenden Endpunkten. Wir können dann eine Aufschneidung der Riemannschen Fläche so bilden, daß die Schnittkurven diese Bahnlinien nicht treffen. Bei Bahnkurven mit gleichen Endpunkten $\lambda$ muß der Schnitt dabei natürlich durch $\lambda$ geführt werden. Auf diese Weise erhalten wir eine durch diese Aufschneidung festgelegte Matrixfunktion, für die $A$ eine Randstelle wird. Diese Betrachtung zeigt, daß die Gesamtheit aller möglichen $f(A)$ bei freier Bewegungsmöglichkeit auf der Riemannschen Fläche identisch wird mit der Gesamtheit aller Werte von $f(A)$ bei allen möglichen Aufschneidungen der Riemannschen Fläche. Unsere Einschränkung ist damit als nicht wesentlich für die Ermittlung des gesamten Wertevorrates einer mehrdeutigen analytischen Matrixfunktion erkannt.

Ist $y = f(x)$ die Umkehrfunktion einer eindeutigen analytischen Funktion $x = \varphi(y)$, so gilt bei $B(t) = f(A(t))$ für die definierende Folge $A(t)$ wegen der Verschiedenheit der Eigenwerte von $A(t) : A(t) = \varphi(B(t))$ und damit $A = \varphi(B)$. $B$ ist also Lösung der Matrixgleichung

$$A = \varphi(B) \tag{4.1}$$

mit bekanntem $A$. Umgekehrt liefert jede Lösung $B$ von (4.1) einen der möglichen Werte von $f(A)$ bei passender Aufschneidung. Für solche $f(x)$ ist also die Aufgabe der Gewinnung aller Werte von $f(A)$ äquivalent zur Aufgabe der Lösung der Matrixgleichung (4.1), in der $B$ eine reguläre Stelle für die Matrixfunktion $\varphi(B)$ darstellt. Aus dieser Äquivalenz folgt in Verbindung mit Satz 5 weiterhin, daß wir uns bei der Lösung von (4.1) auf die $A$ mit lauter zusammenfallenden Eigenwerten beschränken können. Außerdem können wir gemäß § 2 voraussetzen, daß sich $A$ in der Normalform (2.3) befindet, wo jetzt alle $\lambda_k$ gleich $\lambda$ sind.

Zu jeder Umordnung der Kästchen von $A$ gehöre nun eine bestimmte Matrix $U$, die diese Umordnung bewirkt. $T$ bezeichne eine mit $A$ vertauschbare Matrix.

Denken wir uns nun das unbekannte $B$ vermöge der Transformation mit einem $V$ ebenfalls in Normalform (2.3) gebracht,

$$B = V \cdot \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_k \end{pmatrix} V^{-1},$$

wo die Eigenwerte $\mu_\nu$ der $B_\nu$ teilweise gleich sein können, jedoch stets $\varphi(\mu_\nu) = \lambda$ gilt, so ist nach den Ergebnissen des § 3:

$$A = V \cdot \begin{pmatrix} \varphi(B_1) & & 0 \\ & \ddots & \\ 0 & & \varphi(B_k) \end{pmatrix} V^{-1} \tag{4.2}$$

Ist nun beliebig gewählt

$$W = \begin{pmatrix} W_1 & & 0 \\ & \ddots & \\ 0 & & W_k \end{pmatrix},$$

so daß $W_\nu^{-1} \varphi(B_\nu) W_\nu$ in Normalform (2.3) gebracht wird, so ist $VW = TU$ mit passenden $T$ und $U$ und damit

$$A = V \cdot W \begin{pmatrix} W_1^{-1}\,\varphi(B_1)\,W_1 & & 0 \\ & \ddots & \\ 0 & & W_k^{-1}\,\varphi(B_k)\,W_k \end{pmatrix} W^{-1}\,V^{-1} =$$

$$= T\,U \begin{pmatrix} \varphi(B_1') & & 0 \\ & \ddots & \\ 0 & & \varphi(B_k') \end{pmatrix} U^{-1}\,T^{-1}$$

mit

$$B_\nu' = W_\nu^{-1}\,B_\nu\,W_\nu.$$

Wir haben also wegen $A\,T = T\,A$

$$(4.3)\qquad \left\{\begin{aligned} U^{-1} A\,U &= \varphi\begin{pmatrix} B & & 0 \\ & \ddots & \\ 0 & & B_k' \end{pmatrix} \\ \text{und}\qquad & \\ B &= T \cdot U \cdot \begin{pmatrix} B_1' & & 0 \\ & \ddots & \\ 0 & & B_k' \end{pmatrix} U^{-1}\,T^{-1}. \end{aligned}\right.$$

Hierbei sind die $B_\nu'$-Matrizen, deren Normalform (2.3) nur aus einem einzigen Kästchen besteht: irreduzible Matrizen.

Zur Auffindung von $B$ wird also $A$ zunächst durch ein $U^{-1}$ umgeordnet, die Kästchen nach der Umordnung dann in neue Kästchen $A_\nu'$ zusammengefaßt, die als $\varphi(B_\nu')$ mit irreduziblen $B_\nu'$ geschrieben werden können. $B$ erhält man dann durch Rücktransformationen der aus den $B_\nu'$ bestehenden Matrix mit Hilfe von $U$ und zusätzliche Transformation durch eine beliebige mit $A$ vertauschbare Matrix $T$. Dabei ist bis jetzt noch nicht klar, ob zur Auffindung aller $B$ auch alle der endlich vielen $U$ und der endlich vielen Zusammenfassungen zu den $A_\nu'$ statthaft sind. Dagegen ist $T$ sicher völlig beliebig wählbar, da bei $A = \varphi(B)$ auch $A = TAT^{-1} = T\,\varphi(B)\,T^{-1} = \varphi(TBT^{-1})$ ist, so daß mit $B$ auch stets $TBT^{-1}$ eine Lösung darstellt. Es ist durchaus möglich, daß es zu einem der $A_\nu'$ mehrere zugehörige $B_\nu'$ gibt: $B_{\nu 1}', B_{\nu 2}', \ldots$ mit den jeweils zusammenfallenden Eigenwerten $\mu_{\nu 1}, \mu_{\nu 2}, \ldots$. Ist hierbei $\mu_{\nu i} = \mu_{\nu k}$, so sind $B_{\nu i}'$ und $B_{\nu k}'$ wegen der Irreduzibilität Transformierte voneinander: $B_{\nu i}' = SB_{\nu k}'\,S^{-1}$. Es ist dann $A_\nu' = \varphi(B'_{\nu i}) = S\,\varphi(B_{\nu k}')\,S^{-1} = S\,A_\nu'\,S^{-1}$. $A_\nu'$ ist also mit $S$ vertauschbar. Wir können daher annehmen, daß $S$ bereits in $T$ enthalten ist. Die wesentlich verschiedenen Möglichkeiten für $B_\nu'$ müssen also zu lauter verschiedenen Werten für $\mu_\nu = f(\lambda)$ gehören. Hieraus folgt

*Satz 9*: Ist $T$ eine mit $A$ vertauschbare Matrix mit $|T| \neq 0$, so ist mit $B$ auch $TBT^{-1}$ eine Lösung der Matrixgleichung $A = \varphi(B)$ mit eindeutigem analytischen $x = \varphi(y)$. Bis auf Transformation mit allen $T$ ist die Lösung der Matrixgleichung höchstens endlich- oder abzählbar unendlich-deutig, je nachdem dies die Umkehrfunktion $y = f(x)$ von $x = \varphi(y)$ ist.

Es bleibt jetzt nur noch die Frage, wie ein $A'$ aussehen muß, damit es als $\varphi(B')$ mit irreduziblem $B'$ geschrieben werden kann. Dies läßt sich nun leicht klären unter Verwendung der beiden folgenden Hilfssätze:

*Hilfssatz 2*: Es bezeichne $Q_n$ die Matrix vom Grade $n$ mit den Komponenten

$$q_{ik} = \begin{Bmatrix} 1 \text{ für } k - i = 1 \\ 0 \text{ für } k - i \neq 1 \end{Bmatrix}.$$

Dann ist bei $a_l \neq 0$ mit $l \geq 1$ die Matrix

$$P = a_l \cdot Q_n^l + a_{l+1} \cdot Q_n^{l+1} + \cdots + a_{n-1} \cdot Q_n^{n-1}$$

äquivalent zu $Q_n^l$.

*Beweis*: Es seien $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ die Grundvektoren. Führen wir noch $\mathfrak{e}_\nu = 0$ für $\nu \leq 0$ ein, so gilt allgemein $Q_n^l \mathfrak{e}_\nu = \mathfrak{e}_{\nu - l}$ und damit:

$$P\,\mathfrak{e}_\nu = \sum_{\varkappa=0}^{\infty} a_{l+\varkappa}\, \mathfrak{e}_{\nu-l-\varkappa}\,. \tag{4.4}$$

Aus der Funktion $\psi(x) = a_l + a_{l+1}\,x + \cdots$ gewinnen wir für $r = 0, 1, \ldots$ Koeffizienten $b_{r,\mu}$ gemäß

$$\psi^{-r}(x) = \sum_{\mu=0}^{\infty} b_{r,\mu}\, x^\mu\,.$$

Wir bilden nun die Vektoren

$$\mathfrak{f}_s = \sum_{\mu=0}^{\infty} b_{r,\mu}\, \mathfrak{e}_{s-\mu} \qquad (s = 1, \ldots, n)\,,$$

wobei $r$ so bestimmt wird, daß $0 < s - r\,l \leq l$ ist. Die $\mathfrak{f}_s$ sind dann wegen $b_{r,0} \neq 0$ linear unabhängig voneinander, und es gilt wegen (4.4):

a) für $r = 0$, also $s \leq l$: $P \cdot \mathfrak{f}_s = 0$;

b) für $r \geq 1$:

$$P\,\mathfrak{f}_s = \sum_{\mu=0}^{\infty} \sum_{\varkappa=0}^{\infty} b_{r,\mu}\, a_{l+\varkappa}\, \mathfrak{e}_{s-\mu-l-\varkappa} = \sum_{\omega=0}^{\infty} \mathfrak{e}_{s-l-\omega} \cdot \sum_{\mu+\varkappa=\omega} b_{r,\mu}\, a_{l+\varkappa} =$$
$$= \sum_{\omega=0}^{\infty} b_{r-1,\omega}\, \mathfrak{e}_{s-l-\omega} = \mathfrak{f}_{s-l}\,.$$

Wählen wir nun die $\mathfrak{f}_s$ als neue Grundvektoren, so wird $P$ in $Q_n^l$ transformiert.

*Hilfssatz 3*: Die Normalform (2.3) von $Q_n^l$ besteht aus $k$ Kästchen vom Grade $(m+1)$ und $(l-k)$ Kästchen vom Grade $m$, wenn $n = m\,l + k$ mit $k < l$ ist.

*Beweis*: Es ist $Q_n^l \mathfrak{e}_\nu = \mathfrak{e}_{\nu-l}$, wenn wieder $\mathfrak{e}_0 = \mathfrak{e}_{-1} = \cdots = 0$ gesetzt ist. Wählen wir als Transformierende die Matrix, die die $\mathfrak{e}_\nu$ in die folgende Reihenfolge bringt:

$$\left.\begin{array}{l} \mathfrak{e}_1, \mathfrak{e}_{1+l}, \mathfrak{e}_{1+2l}, \ldots \\ \mathfrak{e}_2, \mathfrak{e}_{2+l}, \ldots \\ \ldots\ldots\ldots\ldots \\ \mathfrak{e}_k, \mathfrak{e}_{k+l}, \ldots \end{array}\right\} \text{ je } (m+1) \text{ Stück}$$

$$\left.\begin{array}{l} \mathfrak{e}_{k+1}, \mathfrak{e}_{k+1+l}, \ldots \\ \ldots\ldots\ldots\ldots\ldots \\ \mathfrak{e}_l, \mathfrak{e}_{2l}, \ldots \end{array}\right\} \text{ je } m \text{ Stück}$$

so wird $Q_n$ in die angegebene Normalform transformiert.

*Bemerkung*: Die Hilfssätze 2 und 3 gelten trivialerweise auch bei $l \geq n$, da dann $Q_n^l = 0$ ist; nicht dagegen bei $l = 0$.

Es möge nun $A' = \varphi(B')$ mit irreduziblem $B'$ vom Grade $n$ sein. $B'$ ist also äquivalent zu $\mu E + Q_n$, wo $Q_n$ wie in Hilfssatz 2 definiert ist. Es ist dann $A'$ äquivalent zu $\varphi(\mu E + Q_n)$. Hat nun $\varphi(y)$ an der Stelle $\mu$ die Entwicklung

$$\varphi(y) = \lambda + a_l (y - \mu)^l + a_{l+1} (y - \mu)^{l+1} + \cdots \text{ mit } a_l \neq 0 \text{ bei } l \geq 1,$$

so wird nach Satz 6:

$$\varphi(\mu E + Q_n) = \lambda E + a_l Q_n^l + a_{l+1} Q_n^{l+1} + \cdots,$$

was nach den Hilfssätzen 2 und 3 in der Normalform (2.3) $k$ Kästchen vom Grade $(m + 1)$ und $(l - k)$ Kästchen vom Grade $m$ hat, wenn $n = m\,l + k$ mit $k < l$ ist. Dies ist dann auch die Normalform von $A'$. Hat umgekehrt $A'$ diese Normalform, so geht es gemäß Hilfssatz 3 bei einer Umordnung der Zeilen und Spalten über in $\lambda E + Q_n$. Ist nun $y = f(x) = \mu + b \sqrt[l]{x - \lambda} + \cdots$, so wird, abgesehen von dieser Umordnung, $B' = f(A') = \mu E + b \cdot Q_n + \cdots$ die bis auf Transformation mit einer mit $A'$ vertauschbaren Matrix $T'$ eindeutig bestimmte Lösung der Matrixgleichung $A' = \varphi(B')$ mit irreduziblem $B'$ des Eigenwertes $\mu$. Damit haben wir einen vollkommenen Überblick über die Lösungsmannigfaltigkeit von $B = f(A)$ bei solchen $A$ mit lauter gleichen Eigenwerten $\lambda$ gewonnen:

*Satz 10*: $A$ möge nur den mehrfachen Eigenwert $\lambda$ haben. Es befinde sich in Normalform. Es seien

$$f_\nu(x) = \mu_\nu + b_\nu \cdot \sqrt[l_\nu]{x - \lambda} + \cdots, \qquad \nu = 1, 2, \cdots$$

die verschiedenen Zweige von $f(x)$ an der Stelle $x = \lambda$. Dann ordne man die Kästchen von $A$ mit Hilfe eines $U^{-1}$ so um, daß man mit geeigneten Zahlen $k_\nu$ und $m_\nu$ bei $0 \leq k_\nu < l_\nu$ und $m_\nu \geq 0$ Zusammenfassungen $A'_\nu$ von $k_\nu$ Kästchen vom Grade $(m_\nu + 1)$ mit $(l_\nu - k_\nu)$ Kästchen vom Grade $m_\nu$ bilden kann. Es ist dann $f(A'_\nu) = B'_\nu$ mit irreduziblem $B_\nu$, wenn der Zweig $f_\nu$ verwendet wird. $f(A)$ entsteht dann aus der aus den $B_\nu$ gebildeten Matrix durch Rücktransformation mit $U$ und nachträglicher Transformation mit einer beliebigen mit $A$ vertauschbaren Matrix $T$.

*Bemerkung 10a*: Sind insbesondere alle $l_\nu = 1$, so wird $f(x)$ auf die Kästchen der Normaldarstellung von $A$ einzeln angewendet und liefert für jedes Kästchen soviele Möglichkeiten, wie Zweige von $f(x)$ vorhanden sind. Die Anwendung von $U$ ist in diesem Falle unnötig. $f(A)$ ist stets lösbar.

*Bemerkung 10b*: Ist für ein bestimmtes $A$ die angegebene Vorschrift nicht durchführbar, so ist $B = f(A)$ als Umkehrung von $A = \varphi(B)$ unlösbar.

In dem eingangs genannten *Beispiel* $f(x) = \sqrt{x}$ ist für $\lambda \neq 0 : l_1 = l_2 = 1$. Für ein Kästchen von $A$ der Gestalt $\lambda E + Q$, $E$ und $Q$ vom Grade $n_\nu$, ist:

$$f(\lambda E + Q) = \pm \sqrt{\lambda} \cdot \sum_{\varkappa=0}^{n_\nu - 1} \binom{\frac{1}{2}}{\varkappa} \left(\frac{Q}{\lambda}\right)^\varkappa$$

genau zweideutig. Hat $A$ die Zahl von $z_s$ Kästchen vom Grade $s$, so hat $f(A)$ bis auf Äquivalenz genau $\prod_s (z_s + 1)$ verschiedene Lösungen. Ist dagegen $\lambda = 0$, so ist $l = 2$. Eine Zusammenfassung $A_\nu$ besteht dann (Typus I)

entweder aus 2 Kästchen des gleichen Grades $m_\nu \geq 1$; es ist dann nach Hilfssatz 2 $\sqrt{A_\nu}$ äquivalent zu $Q_{2m_\nu}$. Oder $A'_\nu$ besteht (Typus II) aus 2 Kästchen der Grade $m_\nu + 1$ und $m_\nu \geq 1$; es ist dann $\sqrt{A'_\nu}$ äquivalent zu $Q_{2m_\nu+1}$. Oder $A'_\nu$ ist (Typus III) ein Kästchen vom Grade 1 mit $\sqrt{A'_\nu} = 0$. Nach vorgenommener Zusammenfassung ist hier $f(A)$ bis auf Transformation durch ein mit $A$ vertauschbares $T$ eindeutig bestimmt. Bis auf Äquivalenz ist die Anzahl der Lösungen von $f(A)$ also gleich der Anzahl der Möglichkeiten, die Matrix $A$ in $\alpha_r$ Stück des Typus I mit $m_\nu = r$, $\beta_r$ Stück des Typus II mit $m_\nu = r$ und $\beta_0$ Stück des Typus III zusammenzufassen. Hat nun $A$ wieder $z_s$ Kästchen vom Grade $s$, so ist:

$$z_s = 2\,\alpha_s + \beta_{s-1} + \beta_s \qquad \text{für } s = 1, 2, \ldots, N, \tag{4.5}$$

wo $N$ der größte vorkommende Grad eines Kästchens von $A'$ ist. Dabei ist $\beta_N = 0$ zu setzen. Um die möglichen $A'_\nu$ zu finden, haben wir (4.5) bei gegebenen $z_s$ mit nichtnegativen $\alpha_s$ und $\beta_s$ zu lösen. Nehmen wir an, wir hätten eine solche Lösung, so können wir, falls vorhanden, ein $\beta_s \geq 2$ um 2 vermindern und dafür $\alpha_s$ und $\alpha_{s+1}$ um 1 erhöhen und so eine neue Lösung erhalten. Wenn überhaupt, so gibt es also auch eine Lösung $\alpha'_s$, $\beta'_s$, wo alle $\beta'_s = 0$ oder $= 1$ sind. Aus (4.5) folgt unmittelbar, daß dabei

$$\beta'_s \equiv z_{s+1} + z_{s+2} + \cdots + z_N \pmod 2$$

ist, woraus sich dann $\alpha'_s = \frac{1}{2}(z_s - \beta_{s-1} - \beta'_s)$ ergibt, was automatisch ganzzahlig ist. Ist dabei ein $\alpha'_s < 0$, so ist $\sqrt{A}$ nicht lösbar. Sind dagegen alle $\alpha'_s \geq 0$, so gibt es wenigstens diese eine Lösung, aus der wir alle anderen durch mehrmalige Anwendung des folgenden Prozesses gewinnen: Zwei aufeinanderfolgende $\alpha_\nu \geq 1$ und $\alpha_{\nu+1} \geq 1$ werden beide um 1 vermindert und dafür $\beta_\nu$ um 1 erhöht. Die Anzahl der Lösungen von (4.5) ist dann gleich der Anzahl der nichtäquivalenten Lösungen von $f(A) = \sqrt{A}$. Man bemerke übrigens, daß es stets wenigstens eine Lösung gibt, wenn alle $z_s \geq 1$ sind, da $\beta'_{s-1} + \beta'_s \equiv z_s \pmod 2$ und $\beta'_{s-1} + \beta'_s \leq 2$ und damit sicher $z_s - \beta'_{s-1} - \beta'_s \geq 0$ ist.

Für zweireihige Matrizen zeigen diese Ergebnisse die folgenden Möglichkeiten:

$\lambda \neq 0$: $\quad \sqrt{\begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}} = \pm\sqrt{\lambda}\cdot E,\quad \sqrt{\lambda}\cdot\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$ 3 Lösungen

$\sqrt{\begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix}} = \pm\sqrt{\lambda}\begin{pmatrix}1 & \frac{1}{2\lambda}\\ 0 & 1\end{pmatrix}$ 2 Lösungen

$\lambda = 0$: $\quad \sqrt{\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}}$ mit $\begin{cases}\alpha_1 = 0,\ \beta_0 = 2: = \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}\\ \alpha_1 = 1,\ \beta_0 = 0: = \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\end{cases}$ 2 Lösungen

$\sqrt{\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}}$ mit $\begin{cases}z_1 = 0,\ z_2 = 1;\\ \text{also } \beta'_0 = \beta'_1 = 1;\\ \underline{\alpha'_1 = -1}\end{cases}$ keine Lösung.

### § 5. Differentiale analytischer Matrixfunktionen.

Es sei jetzt wieder $f(A)$ eine eindeutige analytische Matrixfunktion im Sinne von § 3. $A$ sei eine Matrix, für deren Eigenwerte $f(x)$ bei passender Wahl der Aufschneidung der Riemannschen Fläche analytisch ist. Nach Satz 7 ist dann bei $t > 0$:

$$\lim_{t \to 0} f(A + t D) = f(A) \tag{5.1}$$

für beliebiges $D$; insbesondere braucht nicht darauf geachtet zu werden, ob $A + t D$ zusammenfallende Eigenwerte besitzt oder nicht. Es hat daher einen eindeutigen Sinn, die „Variation in der Richtung $D$" einzuführen[5]) gemäß

$$\delta f(A) = f(A + t D) - f(A), \quad t > 0, \quad t \ll 1. \tag{5.2}$$

Nach obiger Voraussetzung existiert auch eindeutig $f'(A)$. Es liegt daher nahe zu fragen, ob man $\delta f(A)$ bis auf höhere Potenzen in $t$ durch $t \cdot f'(A) D$ ersetzen darf. Daß dies nicht allgemein möglich ist, zeigt bereits das Beispiel $f(x) = x^2$, wo

$$\delta f(A) - t f'(A) D = t \cdot (D A - A D) + O(t^2) \tag{5.3a}$$

ist, so daß die Vertauschbarkeit von $D$ mit $A$ zu verlangen wäre. Ist jedoch nun $G$ eine beliebige mit $A$ vertauschbare Matrix, so folgt aus (5.3a) nach Multiplikation mit $G$ und Spurbildung wegen der für Spuren gültigen Rechenregel $\{F_1 F_2 \ldots F_n\} = \{F_n F_1 \ldots F_{n-1}\}$:

$$\{G(\delta f(A) - t f'(A) D)\} = O(t^2) \qquad \text{auch bei } A D \neq D A. \tag{5.3b}$$

Die Gültigkeit einer solchen Beziehung unabhängig von $D$ ist deshalb besonders wichtig, weil in den physikalischen Anwendungen die Invarianten der vorkommenden Tensoren wesentlich sind, die sich ja durch die Spuren ausdrücken lassen. Es ist daher bemerkenswert, daß wir den soeben in einem ganz speziellen Beispiel gefundenen Sachverhalt allgemein aussprechen können im

*Satz 11*: Ist $f(A)$ eine analytische Matrixfunktion und $f(x)$ an den Stellen der Eigenwerte von $A$ regulär, so gilt bei $\delta f(A) = f(A + t D) - f(A)$:

$$\delta f(A) = t \cdot f'(A) D + O(t^2), \qquad \text{falls } A D = D A$$

$$\{G(\delta f(A) - t \cdot f'(A) D)\} = O(t^2), \qquad \text{falls } A G = G A \text{ und } D \text{ beliebig.}$$

Wir beweisen den Satz zunächst für den Spezialfall, daß $A$ lauter gleiche Eigenwerte $\lambda$ hat und führen dann den allgemeinen Fall darauf zurück. Setzen wir $f(x) = g(x - \lambda)$, so wird

$$\delta f(A) = g(A - \lambda E + t D) - g(A - \lambda E) = \delta g(A - \lambda E),$$

$$f'(A) = g'(A - \lambda E)$$

und

$$G \cdot (A - \lambda E) = (A - \lambda E) \cdot G.$$

---

[5]) Auf solche Variationen wird man in den physikalischen Anwendungen zwangsläufig geführt, z. B. wenn man die Änderung des Spannungstensors bei zeitlich variablem Verzerrungszustand studieren will. Vgl. Richter, H.: Das isotrope Elastizitätsgesetz. Z. ang. Math. Mech. 28, 205—209 (1948).

Wir können daher ohne Einschränkung der Allgemeinheit $\lambda = 0$ annehmen. Es sei nun für kleine $x$:

$$f(x) = \sum_{\nu=0}^{\infty} a_\nu x^\nu .$$

Wir setzen dann mit noch zu bestimmendem $N$:

$$f_1(x) = \sum_{\nu=0}^{N-1} a_\nu x^\nu , \; f_2(x) = \sum_{\nu=N}^{\infty} a_\nu x^\nu ;$$

dann ist

$$\delta f(A) - t f'(A) D = [\delta f_1(A) - t f_1'(A) D] + [\delta f_2(A) - t f_2'(A) D],$$

so daß es genügt, den Satz einzeln für $f_1(x)$ und $f_2(x)$ zu beweisen. Wir haben nun

$$\delta f_1(A) - t f_1'(A) D =$$
$$= t \cdot \sum_{\nu=0}^{N-1} a_\nu (A^{\nu-1} D + A^{\nu-2} D A + \cdots + D A^{\nu-1} - \nu A^{\nu-1} D) + O(t^2),$$

woraus die Behauptung des Satzes für $f_1(A)$ unmittelbar folgt.

Schreiben wir weiter gemäß (2.6)

$$f_2(A + tD) = \sum_{k=0}^{n-1} c_k(t) (A + tD)^k \quad \text{und} \quad f_2'(A + tD) = \sum_{k=0}^{n-1} d_k(t) (A + tD)^k,$$

so erhält man aus (3.6) bei genügend großem $N$ wegen $f_2(x) = O(x^N)$:

$$c_k(t) = 0\,(t^2) \quad \text{und} \quad d_k(t) = O(t^2)$$

und damit

$$\delta f_2(A) - t f_2'(A) \cdot D = O(t^2),$$

was die Richtigkeit der Behauptungen für $f_2(A)$ zeigt.

Es möge nun $A$ teilweise verschiedene Eigenwerte haben. Nach einer geeigneten Transformation können wir annehmen, daß $A$ reduziert ist:

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$$

mit eigenwertfremden $A_\nu$. Eine beliebige Matrix $P$ des gleichen Grades wie $A$ schreiben wir dann in der Gestalt

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix},$$

wo $P_{11}$ vom gleichen Grad wie $A_1$ ist, und nennen:

$$(5.4) \qquad \begin{cases} P \text{ reduziert:} & P = (R), \text{ wenn } P_{12} = P_{21} = 0, \\ P \text{ komplementär:} & P = (K), \text{ wenn } P_{11} = P_{22} = 0. \end{cases}$$

Es gelten dann die Rechenregeln:

$$(5.5) \qquad \begin{cases} (R) \cdot (R) = (R); & (K) \cdot (K) = (R) \\ (K) \cdot (R) = (K); & (R) \cdot (K) = (K) \\ \{K\} = 0. \end{cases}$$

Das gegebene $D$ schreiben wir als Summe:

$$(5.6) \qquad D = D_r + D_k \quad \text{mit} \quad D_r = (R) \quad \text{und} \quad D_k = (K).$$

Schließlich ist wegen der Eigenwertfremdheit der $A_\nu$ und der Vertauschbarkeit von $G$ mit $A$:

$$G = (R). \tag{5.7}$$

Ist $D$ mit $A$ vertauschbar, so ist $D_k = 0$. Durch Satz 5 wird unsere Behauptung auf die entsprechende Behauptung für die $A_\nu$ und damit schließlich auf die für $A$ mit lauter gleichen Eigenwerten zurückgeführt. Sei nun $D_k \neq 0$, so schreiben wir für den Beweis der zweiten Behauptung unseres Satzes gemäß (2.6) für kleine $t$ und $u$ unter Beachtung von (5.5):

$$\begin{aligned} f(A + tD_r + uD_k) &= \sum_{\nu=1}^{n-1} c_\nu(t, u)\,(A + tD_r + uD_k)^\nu = \\ &= \sum_{\nu=1}^{n-1} c_\nu(t, u)\,[(A + tD_r)^\nu + (K) + O(u^2)]\,, \end{aligned}$$

oder

$$\begin{aligned} f(A + tD_r + uD_k) &= f(A + tD_r) + \\ &+ \sum_{\nu=1}^{n-1} [c_\nu(t, u) - c_\nu(t, 0)] \cdot (A + tD_r)^\nu + (K) + O(u^2)\,. \end{aligned} \tag{5.8}$$

Bei festem $t$ hängen die $c_\nu(t, u)$ eindeutig analytisch ab von den Potenzsummen

$$s_\varkappa(u) = \{(A + tD_r + uD_k)^\varkappa\} = \{(A + tD_r)^\varkappa + (K) + O(u^2)\} = s_\varkappa(0) + O(u^2)\,.$$

Also ist $c_\nu(t, u) - c_\nu(t, 0) = 0\,(u^2)$, so daß aus (5.8) folgt, wenn wir schließlich speziell $u = t$ setzen;

$$f(A + tD) = f(A + tD_r) + (K) + O(t^2).$$

Wir erhalten dann unter Beachtung von $f'(A) = (R)$:

$$\begin{aligned} \delta f(A) - t \cdot f'(A)\,D &= f(A + tD) - f(A) - t \cdot f'(A)\,D = \\ &= f(A + tD_r) - f(A) - t \cdot f'(A)\,D_r + (K) + O(t^2), \end{aligned}$$

und da für $D = D_r$ die erste Behauptung des Satzes bereits bewiesen ist;

$$\delta f(A) - t \cdot f'(A)\,D = (K) + O(t^2),$$

woraus nach Multiplikation mit $G$ und Spurbildung unter Beachtung von (5.7) und (5.5) die Behauptung folgt.

*(Eingegangen am 13. Mai 1949)*

Math. Annalen, Bd. 122, S. 35—36 (1950).

# Verallgemeinerung einer Minkowskischen Ungleichung über konvexe Körper mit Mittelpunkt.

Von

THEODOR SCHNEIDER in Göttingen.

Ein gegebener nicht-konkaver Körper $\mathfrak{K}$ mit Mittelpunkt, den wir uns im Nullpunkt $O$ eines orthogonalen Koordinatensystems gelegen denken wollen, und vom Volumen $V$, werde in bezug auf seinen Mittelpunkt als Fixpunkt mit dem Faktor $\lambda_1$ so dilatiert, daß der entstehende Körper $\mathfrak{K}_1$ keinen Gitterpunkt außer dem Nullpunkt im Innern enthalte, aber wenigstens einen Gitterpunkt $P_1$ auf dem Rande. Gitterpunkte seien dabei Punkte mit ganzzahligen Werten für die Koordinaten. Der Vektor $\overrightarrow{OP_1}$ sei $\mathfrak{p}_1$ genannt. Der Vektor $\mathfrak{p}_2$ führe von $O$ nach dem Gitterpunkt $P_2$, der nicht auf der Geraden $O\,P_1$, aber auf dem Rande von $\mathfrak{K}_2$, das aus $\mathfrak{K}$ durch Dilatation mit $\lambda_2$ hervorgehe, liege, und es enthalte $\mathfrak{K}_2$ außerhalb der Geraden durch $O$, $P_1$ keine Gitterpunkte im Innern. Es sind daher $\mathfrak{p}_1$ und $\mathfrak{p}_2$ linear unabhängig und ferner ist $\lambda_1 \leq \lambda_2$. So fortfahrend bestimme man $P_3$, das nicht in der Ebene $O\,P_1\,P_2$, aber auf dem Rand von $\mathfrak{K}_3$ liegt, wobei $\mathfrak{K}_3$ aus $\mathfrak{K}$ durch Dilatation mit $\lambda_3$ hervorgeht und außerhalb der Ebene durch $O$, $P_1$, $P_2$ keine Gitterpunkte im Innern enthält. Durch sinngemäße Weiterführung dieser Bestimmungen erhält man so ein System linear unabhängiger Vektoren $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$, wenn $n$ die Dimension von $\mathfrak{K}$ bedeute, und man erhält außerdem die Dilatationskoeffizienten $\lambda_1, \ldots, \lambda_n$ mit $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Dann lautet die Minkowskische Ungleichung über konvexe Körper mit Mittelpunkt:

$$\lambda_1 \ldots \lambda_n \cdot V \leq 2^n.$$

Für diese Ungleichung gibt MINKOWSKI in seinem Buch, Geometrie der Zahlen, einen langen, ausführlichen Beweis. H. DAVENPORT teilte einen einfacheren Beweis mit [1]) und vor kurzem fand H. SCHÜLER[2]) eine schöne, noch wesentlich vereinfachte Beweisdarstellung. Der Schülersche Beweis, der indirekt geführt ist, wird vielleicht noch etwas eleganter, wenn man ihn, wie es im folgenden geschehen soll, als direkten Beweis ansetzt. Dabei soll gleichzeitig die nachstehende Verallgemeinerung der Minkowskischen Ungleichung gezeigt werden.

Aus $\mathfrak{K}$ entstehe durch Dilatation mit einem geeigneten Faktor $\mu_1$ der Körper $\mathfrak{K}_1$, der außer dem Nullpunkt $O$ und $m$ Paaren von symmetrisch zu $O$ gelegenen Gitterpunkten $Q_1^+, Q_1^-; \ldots; Q_m^+, Q_m^-$ keine weiteren Gitterpunkte im Innern, aber mindestens einen auf seinem Rande besitzt. Einer dieser Randpunkte sei $P_1$ und $\overrightarrow{OP_1}$ sei $\mathfrak{p}_1$ genannt. $\mathfrak{K}_2$, das aus $\mathfrak{K}$ durch Dilatation mit dem Faktor $\mu_2$ entstehe, enthalte im Innern außer $O$; $Q_1^+, Q_1^-; \ldots; Q_m^+, Q_m^-$ höchstens Gitterpunkte, die sich auf der Geraden $O\,P_1$ befinden, und mindestens ein Gitterpunkt $P_2$ außerhalb dieser Geraden existiere auf

[1]) DAVENPORT, H.: Quart. J. Math., Oxford Ser. 10, 119 (1939).

[2]) SCHÜLER, H.: Arch. Math. (z. Z. im Druck).

dem Rande. Sinngemäß seien entsprechend $\mu_3, \ldots, \mu_n$ erklärt. Dann lautet, wenn das Volumen von $\mathfrak{K}$ wieder mit $V$ bezeichnet werde, die Behauptung:

$$\mu_1 \ldots \mu_n \cdot V \leq m \cdot 2^n.$$

Beweis: Die Variabeln $y_1, \ldots, y_n$, bezogen auf das Parallelkoordinatensystem der Vektoren $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$, mögen $\mathfrak{K}$ beschreiben. Dann ist $\mathfrak{K}^*$, das die Gesamtheit der Punkte $(\mu_1 y_1, \ldots, \mu_n y_n)$ umfasse, wenn $(y_1, \ldots, y_n)$ Punkt von $\mathfrak{K}$ sei, ebenfalls ein nichtkonkaver Körper mit Mittelpunkt, und sein Volumen ist

$$V^* = \mu_1 \ldots \mu_n \cdot V.$$

Mit $\mathfrak{K}^{(j)}$ sei die Gesamtheit der Punkte

$$(\mu_1 y_1, \mu_2 y_2, \ldots, \mu_{j-1} y_{j-1}, \mu_j y_j, \mu_j y_{j+1}, \ldots, \mu_j y_n)$$

für $j = 1, \ldots, n$ bezeichnet. Dann enthält $\mathfrak{K}^{(1)}$ nach Voraussetzung im Innern keine Gitterpunkte außer $O$; $Q_1^+, Q_1^-; \ldots; Q_m^+, Q_m^-$. Es werde nun die Induktionsvoraussetzung gemacht, $\mathfrak{K}^{(j)}$ enthalte im Innern keine Gitterpunkte außer $O$; $Q_1^+, Q_1^-; \ldots; Q_m^+, Q_m^-$, und es soll gezeigt werden, daß dann im Innern von $\mathfrak{K}^{(j+1)}$ ebenfalls keine anderen Gitterpunkte als $O$; $Q_1^+, Q_1^-; \ldots; Q_m^+, Q_m^-$ vorkommen können. Die Koordinaten in Richtung von $\mathfrak{p}_1, \ldots, \mathfrak{p}_j$ sind beim Übergang von $\mathfrak{K}^{(j)}$ zu $\mathfrak{K}^{(j+1)}$ invariant, so daß keine Gitterpunkte auf der durch die Vektoren $\mathfrak{p}_1, \ldots, \mathfrak{p}_j$ aufgespannten $j$-dimensionalen Hyperebene hinzukommen können. $\mathfrak{K}^{(j+1)}$ ist wegen $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_j \leq \mu_{j+1}$ enthalten in $\mathfrak{K}_{j+1}$, das die Gesamtheit der Punkte $(\mu_{j+1} y_1, \ldots, \mu_{j+1} y_n)$ umfaßt, und der Körper $\mathfrak{K}_{j+1}$ besitzt nach Definition von $\mu_{j+1}$ außerhalb der durch $\mathfrak{p}_1, \ldots, \mathfrak{p}_j$ aufgespannten Hyperebene gewiß keine Gitterpunkte im Innern mit Ausnahme von $O$; $Q_1^+, Q_1^-; \ldots; Q_m^+, Q_m^-$. Also kann auch $\mathfrak{K}^{(j+1)}$ keine neuen Gitterpunkte im Innern besitzen. Durch vollständige Induktion folgt dann, daß $\mathfrak{K}^{(n)}$, das mit $\mathfrak{K}^*$ identisch ist, außer $O$; $Q_1^+, Q_1^-; \ldots; Q_m^+, Q_m^-$ keine Gitterpunkte im Innern enthält.

Nach VAN DER CORPUTS[3]) Verallgemeinerung des Minkowskischen Satzes über nicht-konkave Körper mit Mittelpunkt, angewandt auf den Körper $\mathfrak{K}^*$, folgt nun

$$V^* \leq m \cdot 2^n,$$

und daraus ergibt sich die behauptete Ungleichung:

$$\mu_1 \ldots \mu_n \cdot V \leq m \cdot 2^n.$$

---

[3]) CORPUT, J. G. VAN DER: Acta arithmet. 2, 145 (1936).

*(Eingegangen am 30. Mai 1949.)*

Math. Annalen, Bd. 122, S. 37—46 (1950).

# Über den Einfluß algebraischer Windungspunkte auf die Wachstumsordnung.

Von

Hans Wittich in Karlsruhe.

In mehreren Vorträgen hat E. Ullrich[1]) die Frage nach dem Einfluß algebraischer Windungspunkte auf die Wachstumsordnung gestellt und behandelt. Im folgenden sollen einige Flächen betrachtet werden, die wegen ihres Zusammenhanges mit dieser Frage von einigem Interesse sein dürften. Es handelt sich um einfach zusammenhängende, offene Riemannsche Flächen mit unendlich vielen algebraischen Windungspunkten über drei Grundpunkten. Die Flächen gehören zum parabolischen Typus, so daß ihre Erzeugenden $w = w(z)$ in $|z| < \infty$ eindeutige analytische Funktionen sind. Die Flächen sind von so einfacher Bauart, daß die Untersuchung des asymptotischen Verhaltens von $w = w(z)$ für $|z| \to \infty$ von der Fläche aus möglich ist; sie stützt sich auf Überlegungen von L. Ahlfors, die in den letzten Jahren von mehreren Verfassern[2]) weiterentwickelt worden sind.

1. Im Anschluß an P. J. Myrberg[3]) wird die folgende Klasse Riemannscher Flächen betrachtet.

In $\mathfrak{I}\,\zeta > 0$, $\mathfrak{R}\,\zeta > 0$ sei $K_0, K_1, \ldots$ eine unendliche Folge von Modulkreisen; $K_0$ treffe die positiv reelle Achse in $P_0$, $P_1$, $K_1$ in $P_1$, $P_2, \ldots K_n$ in $P_n, P_{n+1}, \ldots, P_0 = 0$. Weiter gelte monoton $\lim\limits_{n\to\infty} P_n = \infty$. Der Folge $K_n$ wird in $\mathfrak{I}\,\zeta > 0$, $\mathfrak{R}\,\zeta < 0$ eine Folge $K'_n$ von Modulkreisen zugeordnet, die durch Spiegelung der $K_n$ an $\mathfrak{R}\,\zeta = 0$ entstehen; die $P_n$ entsprechenden Punkte seien $P'_n$, wobei also $P'_n$, $P'_{n+1}$ zu $K'_n$ gehören, $P'_0 = P_0$. Bei Identifizierung zugeordneter Punkte auf $K_n$ und $K'_n$ bildet die elliptische Modulfunktion $w = \lambda(\zeta)$ den konstruierten Bereich $B$ auf eine einfach zusammenhängende, unendlich vielblättrige Riemannsche Fläche $\mathfrak{W}$ ab, deren Verzweigungspunkte über den Grundpunkten $w = 0, 1, \infty$ liegen. Der in $\mathfrak{R}\,\zeta > 0$ gelegene Teil von $B$ sei $B_1$. Nach dem Riemannschen Abbildungssatz gibt es eine Funktion $z = z(\zeta)$, die $B_1$ in $\mathfrak{I}\,z > 0$ abbildet, wobei $\zeta = 0$ der Stelle $z = 0$ entspricht und die positiv imaginäre $\zeta$-Achse in die negativ reelle $z$-Achse übergeht. Da die Voraussetzungen des Schwarzschen Spiegelungsprinzips erfüllt sind, leistet die nach diesem Prinzip in $B'_1$ analytisch fortgesetzte Funktion $z = z(\zeta)$ die eineindeutige und konforme Abbildung von $B$ in $|z| < \infty$. Die eindeutige analytische Funktion $w = \lambda(\zeta(z)) = f(z)$ erzeugt als Bild von $|z| < \infty$ die Fläche $\mathfrak{W}$, so daß also $\mathfrak{W}$ vom parabolischen Typus ist. Aus der Art, wie $B$

---

[1]) Ullrich, E.: a) Jber. dtsch. math. Ver. **46** (1936); b) Flächenbau und Wertverteilung. Congrès des Mathématiques à Helsingfors 1938.

[2a]) Ahlfors, L.: Acta math. **58** (1932).

[2b]) Teichmüller, O.: Dtsch. Math. **3** (1938).

[2c]) Wittich, H.: Math. Z. **51** (1948).

[2d]) Wittich, H.: Math. Z. **51** (1948).

[3]) Myrberg, P. J.: Über die Bestimmung des Typus einer Riemannschen Fläche. Ann. Acad. Sci. fenn. Sér. A. Tom XLV, No. 3.

aus Fundamentalgebieten der zum Legendreschen Integral gehörigen Modulfunktion zusammengesetzt ist, erkennt man den Aufbau von $\mathfrak{W}$ aus Halbblättern und kann damit auch den zu $\mathfrak{W}$ gehörigen Streckenkomplex konstruieren. Für $P_n = n$, $P'_n = -n$, $n = 0, 1, \ldots$ erhält man den nachstehenden Streckenkomplex (Fig. 1a—1c). Berühren sich im Punkte $P_n$ $g_n$ Modulkreise, so entspricht einer Umgebung dieses Punktes (Vollumgebung wegen der Identifizierungsvorschrift) durch $w = \lambda(\zeta)$ eine Umgebung eines $g_n$-blättrigen Windungspunktes über dem Grundpunkt $w = \lambda(P_n) = \lambda(P'_n) = (0, 1$ oder $\infty)$. Da sich die Punkte $P_n$ nur in $\zeta = \infty$ häufen, liegen über $w = 0$ und 1

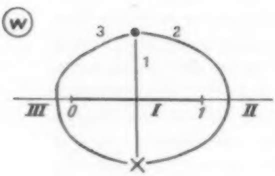

Fig. 1a.

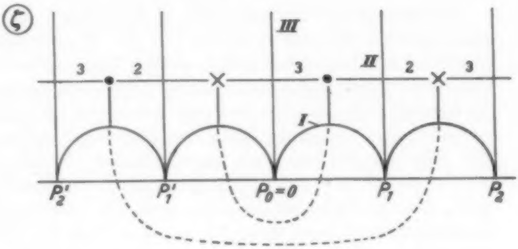

Fig. 1b.

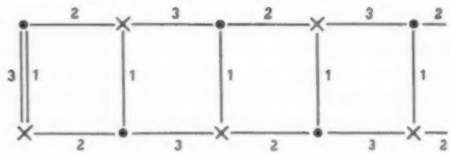
Fig. 1c.

nur algebraische Windungspunkte; über $w = \infty$ liegt genau ein logarithmischer Windungspunkt und möglicherweise algebraische Windungspunkte. Läßt man die den Punkten $P_n$ zugeordneten natürlichen Zahlen $g_n$ bei $n \to \infty$ gegen $\infty$ streben, so erhält man Flächen mit algebraischen Windungspunkten von beliebig hoher Ordnung, eine Tatsache, die bei der eingangs erwähnten Frage besonders interessiert.

Zur Diskussion der solche Flächen erzeugenden gebrochenen Funktionen $w = f(z)$ braucht man hinreichend genaue Aussagen über die Verzerrung bei der Abbildung $\zeta \leftrightarrow z$.

2. Der auf $B$ gelegene Teil von $|\zeta| = \varrho$ sei $K_\varrho$. Das $z$-Bild von $K_\varrho$ ist eine einfache geschlossene Kurve $\Gamma_\varrho$, die $z = 0$ umschlingt. Es sei weiter $\varrho_0 < \varrho_1 < \varrho_2$, $\varrho_0$ passend gewählt. Die Kurven $\Gamma_{\varrho_1}, \Gamma_{\varrho_2}$ bestimmen ein Ringgebiet mit dem Modul $M(\varrho_1, \varrho_2)$, das durch $\tilde z = \tilde z(z)$ in das Rechteck $0 \le \tilde x \le M$, $0 \le \tilde y \le 2\pi$ abgebildet wird, wobei $\Gamma_{\varrho_1}$ in $\tilde z = i\tilde y$ und $\Gamma_{\varrho_2}$ in $\tilde z = M + i\tilde y$ übergehen soll. Weiter sei $\tilde\zeta = \log\zeta$ und $\varrho_1 \le \varrho \le \varrho_2$. Da die Bildkurve von $K_\varrho$ im $\tilde z$-Rechteck eine Länge $\ge 2\pi$ hat, gilt $2\pi \le \int\limits_{\tilde K_\varrho} \left|\frac{d\tilde z}{d\tilde\zeta}\right| |d\tilde\zeta|$;

daraus folgt nach der Schwarzschen Ungleichung und Integration nach $\log\varrho$ zwischen den Grenzen $\varrho_1$ und $\varrho_2$

$$4\pi \log \frac{\varrho_2}{\varrho_1} \leq \int_{K_\varrho} \int_{\varrho_1}^{\varrho_2} \boxed{d\tilde{z}} \leq 2\pi M(\varrho_1, \varrho_2),$$

also

$$2 \log \frac{\varrho_2}{\varrho_1} \leq M(\varrho_1, \varrho_2).$$

Im Rechteck wird nun die Strecke $\tilde{x} = \text{const}$ betrachtet. Bild ist in der $\zeta^* = \log\left(\zeta - \frac{i}{2}\right)$-Ebene eine Kurve der Länge $\geq \pi$. Danach gilt

$$\pi \leq \int_{\tilde{x}} \left|\frac{d\zeta^*}{d\tilde{z}}\right| d\tilde{y}, \quad \pi^2 \leq 2\pi \int_0^{2\pi} \left|\frac{d\zeta^*}{d\tilde{z}}\right|^2 d\tilde{y}$$

oder

$$\frac{\pi}{2}\int_0^M d\tilde{x} = \frac{\pi}{2} M(\varrho_1, \varrho_2) \leq \int_0^M \int_0^{2\pi} \left|\frac{d\zeta^*}{d\tilde{z}}\right|^2 d\tilde{x}\, d\tilde{y} = \int\int \boxed{d\zeta^*} = F.$$

Durch elementare Rechnung findet man für $F$ die für alle $\varrho_1 > e$ gültige Abschätzung

$$F < \pi\left(\log \frac{\varrho_2}{\varrho_1} + \frac{6}{\varrho_1}\right).$$

Mithin gilt für alle $e < \varrho_0 < \varrho_1 < \varrho_2$

$$2 \log \frac{\varrho_2}{\varrho_1} < M(\varrho_1, \varrho_2) < 2 \log \frac{\varrho_2}{\varrho_1} + \frac{12}{\varrho_1}$$

oder

$$\left| M(\varrho_1, \varrho_2) - 2 \log \frac{\varrho_2}{\varrho_1} \right| < \frac{12}{\varrho_1}. \tag{1}$$

Aus (1) folgt mit $r_1(\varrho) = \underset{\Gamma_\varrho}{\mathrm{Min}}\, |z|$, $r_2(\varrho) = \underset{\Gamma_\varrho}{\mathrm{Max}}\, |z|$ nach dem Modulsatz[4]) die Existenz einer endlichen Zahl $\varkappa$ derart, daß $\lim\limits_{\varrho\to\infty} (\log r_1(\varrho) - \log \varrho^2) = \lim\limits_{\varrho\to\infty} (\log r_2(\varrho) - \log \varrho^2) = \varkappa$ gilt, d. h.

$$\left\{\begin{array}{l} \lim\limits_{\varrho\to\infty} \dfrac{r_j(\varrho)}{\varrho^2} = e^\varkappa, \quad j = 1, 2 \quad \text{oder} \\ r_j(\varrho) = e^\varkappa \varrho^2 (1 + \varepsilon_j(\varrho)) \text{ mit } \lim\limits_{\varrho\to\infty} \varepsilon_j(\varrho) = 0 \text{ für } j = 1, 2. \end{array}\right. \tag{2}$$

Diese Verzerrungsaussage bei der Abbildung $\zeta \leftrightarrow z$ ermöglicht die Diskussion der durch die Flächen definierten gebrochenen Funktionen. Es bezeichne $\nu(\varrho, a)$ die Zahl der $a$-Stellen von $w = \lambda(\zeta)$ auf dem von $K_\varrho$ begrenzten endlichen Teil von $B$, wobei jede $a$-Stelle entsprechend ihrer Vielfachheit gezählt wird. Bei der Anzahlfunktion $\nu_1(\varrho, a)$ werde jede $k$-fache $a$-Stelle nur $(k-1)$ mal gezählt. Diese Anzahlfunktionen lassen sich aus der Art, wie $B$ durch die Fundamentalgebiete von $\lambda(\zeta)$ zerlegt wird, bestimmen (und damit auch aus dem Streckenkomplex). Da aus (2)

$$\varrho = A\sqrt{r_1}(1 + \Theta_1(r_1)) = A\sqrt{r_2}(1 + \Theta_2(r_2)), A = e^{-\varkappa/2}, \lim_{r_j\to\infty} \Theta_j(r_j) = 0, \tag{2'}$$

---

[4]) l. c. [2b]).

folgt, bekommt man bei bekannten $\nu(\varrho, a)$, $\nu(\varrho) = \operatorname*{Max}_{a} \nu(\varrho, a)$, $\nu_1(\varrho, a)$ asymptotische Ausdrücke für die in der Wertverteilungslehre gebräuchlichen Anzahlfunktionen $n(r, a)$, $n(r)$, $n_1(r, a)$ bzw. $N(r, a)$, $N(r)$, $N_1(r, a)$.

Für den Streckenkomplex in Fig. 2 findet man

$$n(r, 0) = 2A\sqrt{r}\,(1 + \varepsilon(r)), \qquad n_1(r, 0) = \frac{3}{2}A\sqrt{r}\,(1 + \varepsilon(r)),$$

$$n(r, 1) = 2A\sqrt{r}\,(1 + \varepsilon(r)), \qquad n_1(r, 1) = \frac{3}{2}A\sqrt{r}\,(1 + \varepsilon(r)),$$

$$n(r, \infty) = 2A\sqrt{r}\,(1 + \varepsilon(r)), \qquad n_1(r, \infty) = 0,$$

$$n(r) = 2A\sqrt{r}\,(1 + \varepsilon(r)).$$

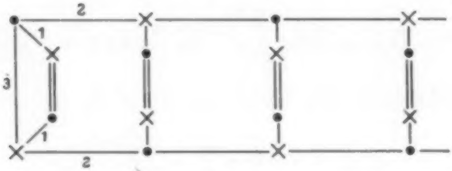


Fig. 2.

Aus diesen Angaben folgt wegen

$$\text{Ordnung } s = \lim_{r\to\infty} \frac{\log n(r)}{\log r}, \quad \delta(a) = 1 - \lim_{r\to\infty} \frac{n(r, a)}{n(r)}, \quad \vartheta(a) = \lim_{r\to\infty} \frac{n_1(r, a)}{n(r)}:$$

$$\delta(\infty) = \frac{1}{2}, \quad \delta(a) = 0 \ \text{ für alle } a \neq \infty,$$

$$\vartheta(0) = \vartheta(1) = \frac{1}{2}, \ \vartheta(a) = 0 \ \text{ für alle } a \neq 0, 1 \text{ und } s = \frac{1}{2},$$

ein Resultat, das von den Funktionen, die durch Flächen mit endlich vielen periodischen Enden definiert werden, bekannt ist[5]).

3. Bei dem in 2. behandelten Beispiel fällt die Wachstumsordnung $s$ mit der durch den Randstellensatz von Ahlfors-Denjoy bestimmten unteren Schranke der Wachstumsordnung zusammen, nämlich $s = \frac{1}{2}$. Dieser Fall liegt stets vor, falls $g_n \leq g$ für alle $n$ gilt. Sind alle algebraischen Elementargebiete so angeordet, daß eine Fläche mit einem periodischen Ende vorliegt, so kann wegen des symmetrischen Aufbaus der Fläche keine Wachstumserhöhung eintreten. Im allgemeinen Falle sieht man die Behauptung etwa so ein. Ist $P_n$ ein beliebiger Modulpunkt auf der positiv reellen Achse, so berühren sich in $P_n$ höchstens $g$ Modulkreise. Dann muß also der Abstand $P_{n-1}\,P_n$ und $P_n\,P_{n+1}$ stets $\geq p > 0$ sein, wobei $p$ allein von $g$, nicht aber von $n$ abhängt. Daraus folgt aber, daß $\nu(\varrho + 1) - \nu(\varrho)$ stets $\leq K$ ist oder $C_1\sqrt{r} \leq n(r) \leq C_2\sqrt{r}$. Wegen $m(r, a) = 0(1)$ für $a \neq 0, 1, \infty$ erhält man also

$$K_1\sqrt{r} < T(r, w) = N(r, a) + m(r, a) + 0(1) < K_2\sqrt{r},$$

[5]) Wittich, H.: Arch. Math. 1 (1948).

d. h.

$$s = \overline{\lim_{r\to\infty}} \frac{\log T(r, w)}{\log r} = \frac{1}{2}.$$

Zu dem gleichen Ergebnis kommt man, wenn die $g_n$ zwar nicht mehr gleichmäßig beschränkt sind, aber hinreichend langsam mit $n$ gegen $\infty$ streben.

Bei dem Streckenkomplex in Fig. 3a werden in die algebraischen Elementargebiete $A_2, A_4, A_6, \ldots$ je $2\,m$ Vierecke eingebaut in der Art, wie es die Figur 3b für $m = 1$ zeigt.

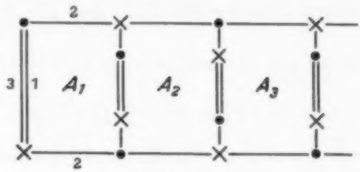


Fig. 3a.

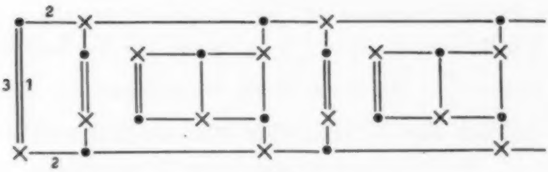


Fig. 3b.

Für diesen Streckenkomplex findet man[6])

$$s = \frac{1}{2}, \quad \delta(0) = \delta(1) = 0, \quad \delta(\infty) = 1 - \frac{2m+2}{2m+4} = \frac{2}{2m+4},$$

$$\vartheta(0) = 1 - \frac{1}{2m+4}, \quad \vartheta(1) = \frac{1}{2} + \frac{1}{2m+4}, \quad \vartheta(\infty) = \frac{1}{2} - \frac{2}{2m+4}.$$

Ordnet man nun den Elementargebieten $A_{2\mu}$ die Zahlen $2\,m_\mu$ zu, wobei $m_\mu \to \infty$ bei $\mu \to \infty$ gilt, so wird man die folgende Verteilung der Defekte und Indices erwarten:

$$\delta(0) = \delta(1) \quad = \delta(\infty) = 0$$

$$\vartheta(0) = 1, \vartheta(1) = \vartheta(\infty) = \frac{1}{2}.$$

Die Wachstumsordnung wird dabei möglicherweise erhöht. Ist $\varrho$ beliebig gegeben, so bestimme man dazu die natürliche Zahl $\mu$ so, daß gilt $2\,\mu + 1 \leq \varrho < 2\,\mu + 3$. Aus dem Streckenkomplex findet man:

---

[6]) l. c. [5]) Formel (7) und (8).

*Zusatz:* In [5]) ist auf S. 165/166 zu setzen:

$\sum_1^4 g_1^{(\mu)}(a_3)\,e_\mu = \frac{\sqrt{2}}{2}$, $\sum_1^4 g_1^{(\mu)}(a_4)\,e_\mu = \frac{3+\sqrt{2}}{2}$, $\vartheta(a_3) = \frac{\sqrt{2}}{7+4\sqrt{2}}$, $\vartheta(a_4) = \frac{\sqrt{2}+3}{7+4\sqrt{2}}$.

$$\nu(\varrho, a) = 4\,\mu + 2\,S(\mu) + 0\,(1) \text{ für alle } a \neq \infty,$$
$$\nu(\varrho, \infty) = 2\,\mu + 2\,S(\mu) + 0\,(1),$$
$$\nu_1(\varrho, 0) = 3\,\mu + 2\,S(\mu) + 0\,(1),$$
$$\nu_1(\varrho, 1) = 3\,\mu + S(\mu) + 0\,(1), \quad \nu_1(\varrho, \infty) = S(\mu) + 0\,(1)$$

mit

$$S(\mu) = \sum_{\varkappa=1}^{\mu} m_\varkappa .$$

Ist $m_\mu = \mu$, so gilt wegen $S(\mu) = \frac{\mu(\mu+1)}{2}$ und $\mu = \frac{\varrho}{2}\left(1 + 0\left(\frac{1}{\varrho}\right)\right)$

$$\nu(\varrho, a) = \frac{\varrho^2}{4}(1 + \varepsilon(\varrho)) \text{ für alle } a, \; \nu_1(\varrho, 0) = \frac{\varrho^2}{4}(1 + \varepsilon(\varrho)) \quad \text{und}$$

$$\nu_1(\varrho, 1) = \nu_1(\varrho, \infty) = \frac{\varrho^2}{8}(1 + \varepsilon(\varrho)),$$

wobei die $\varepsilon(\varrho)$, ebenso wie früher, für die verschiedenen Anzahlfunktionen i. a. verschieden ausfallen, aber alle der Bedingung $\lim\limits_{\varrho\to\infty} \varepsilon(\varrho) = 0$ genügen. Aus diesen Formeln folgt nach (2′) $n(r, a) = r\,\alpha(r)$ für alle $a$, $n_1(r, 0) = r\,\alpha(r)$; $n_1(r, 1) = n_1(r, \infty) = \frac{r}{2}\,\alpha(r)$, wobei für die Funktionen $\alpha(r)$ gilt: $\lim\limits_{r\to\infty} \alpha(r) = \frac{A^2}{4}$. Die durch diesen Streckenkomplex definierte gebrochene Funktion zeigt die angegebene Wertverteilung $\vartheta(0) = 1$, $\vartheta(1) = \vartheta(\infty) = \frac{1}{2}$. Die Ordnung $s$ ist von $\frac{1}{2}$ auf 1 angestiegen. Setzt man $m_\mu = \mu^2$ bzw. $m_\mu = \mu^3$, so erhält man gebrochene Funktionen mit derselben Verteilung der Defekte und Indizes. Die Wachstumsordnungen sind $s = \frac{3}{2}$ bzw. $s = 2$.

Es soll nun noch auf ein Beispiel hingewiesen werden, das zeigt, daß bei schwachem Anwachsen der $m_\mu$ die Ordnung $s = \frac{1}{2}$ nicht erhöht wird.

Es sei $m_1 = 1$, $m_2 = m_3 = 2$, $m_4 = m_5 = m_6 = m_7 = 3, \ldots$, also für $2^\alpha \leq \mu \leq 2^{\alpha+1} - 1$ $m_\mu = \alpha + 1$, $\alpha = 0, 1, \ldots$. Für $\mu = 2^{\alpha+1} - 1$ gilt die Summationsformel $S(\mu) = \alpha\,2^{\alpha+1} + 1$ und für beliebiges $\mu$:

$S(\mu) = (\alpha - 1)\,2^\alpha + (\lambda + 1)(\alpha + 1) + 1$ mit $\mu = 2^\alpha + \lambda$, $\lambda = 0, 1, \ldots, 2^\alpha - 1$.

Wegen $\alpha = \frac{\log \mu}{\log 2}(1 + \varepsilon(\mu))$ erhält man schließlich

$$S(\mu) = \frac{\mu \log \mu}{\log 2}(1 + \varepsilon(\mu))$$

und daraus weiter $\nu(\varrho) = \frac{\varrho \log \varrho}{\log 2}(1 + \varepsilon(\varrho))$ und $n(r) = \frac{A\sqrt{r}}{\log 2} \log \sqrt{r}\,(1 + \varepsilon(r))$, woraus die Behauptung $s = \frac{1}{2}$ folgt. Die Verteilung der Defekte und Indizes ist dieselbe wie bei den anderen Beispielen.

Zufolge des einfachen Flächenbaues läßt sich bei den angegebenen Beispielen verfolgen, wie durch Einschaltung algebraischer Elementargebiete

in den Ausgangskomplex eine Wachstumserhöhung bewirkt werden kann. Die Beispiele sind so gewählt, daß die durch die Flächen definierten gebrochenen Funktionen bei $w = 0$ den Höchstwert 1 für den Index der algebraischen Verzweigtheit aufweisen: $\vartheta(0) = 1$. In den Flächen liegen über $w = \infty$ neben dem einzigen logarithmischen Windungspunkt noch unendlich viele zweiblättrige Windungspunkte und zwar in solcher Häufigkeit, daß der logarithmische Windungspunkt keinen Einfluß mehr auf die Defektverteilung hat. In den konstruierten eindeutigen analytischen Funktionen liegen also solche gebrochenen Funktionen von endlicher Ordnung vor, die

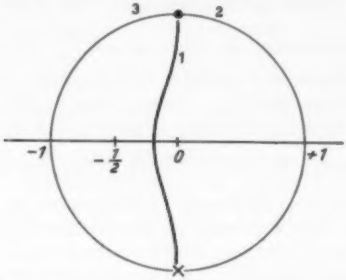


Fig. 4a.

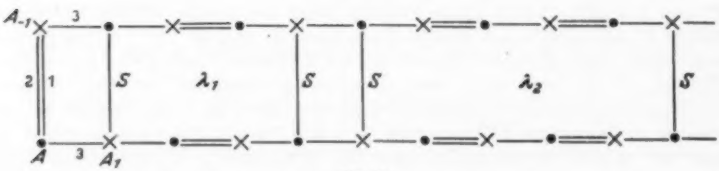


Fig. 4b.

keinen Nevanlinnaschen Ausnahmewert haben, obwohl die von ihnen erzeugten Riemannschen Flächen einen logarithmischen Windungspunkt aufweisen. Das letzte Beispiel ($m_\mu = \alpha + 1$) steht in Zusammenhang mit von G. af Hällström[7]) behandelten Beispielen ganzer Funktionen mit algebraischem Höchstindex. Genau wie bei Hällström wird auch hier durch das Einschalten algebraischer Windungspunkte lediglich der Typus, nicht aber die Ordnung erhöht.

Schaltet man auch in $A_1, A_3, \ldots$ algebraische Elementargebiete ein, so läßt sich i. a. $\vartheta(0) = 1$ nicht mehr erreichen. So gilt, falls man in $A_1$ und $A_2$ je zwei algebraische Elementargebiete einschaltet, in $A_3$ und $A_4$ je drei usw.

$$s = 1, \quad \delta(0) = \delta(1) = \delta(\infty) = 0, \quad \vartheta(0) = \vartheta(1) = \frac{3}{4}, \quad \vartheta(\infty) = \frac{1}{2}.$$

In diesem Zusammenhang soll noch eine von E. Ullrich[8]) behandelte Fläche betrachtet werden. Sie sei durch den obigen Streckenkomplex und Zerschneidungskurve definiert (Fig. 4).

[7]) af Hällström, G.: Math. Z. 47 (1941).
[8]) l. c. [1]).

Die Grundpunkte seien $w = 0, -\frac{1}{2}$ und $\infty$, die Zerschneidungskurve die reelle Achse. Wird die Fläche $\mathfrak{W}$ über $|w| = 1$ zerschnitten, so zerfällt sie in zwei unendlich vielblättrige Flächenstücke $\mathfrak{W}_1$ und $\mathfrak{W}_2$: $\mathfrak{W}_1$ = logarithmisches Windungsflächenstück mit dem logarithmischen Windungspunkt über $w = \infty$, $\mathfrak{W}_2$ enthält unendlich viele algebraische Windungspunkte, deren Ordnungen nicht beschränkt sind. Bringt man auf $\mathfrak{W}_2$ noch Schnitte längs Kurven an, die über der Kurve 1 liegen und den Gliedern $S$ im Streckenkomplex entsprechen, dann zerfällt $\mathfrak{W}_2$ in unendlich viele einfache algebraische Windungsflächenstücke. Die über $w = -\frac{1}{2}$ gelegenen sind durchweg zweiblättrig, die über $w = 0 \frac{\lambda_\varkappa}{2}$-blättrig. Beide Flächenstücke $\mathfrak{W}_1$ und $\mathfrak{W}_2$ lassen sich schlicht und konform bzw. quasikonform auf einfache Normalgebiete abbilden. Durch eine lineare Transformation von $\log w$ geht $\mathfrak{W}_1$ über in $\mathfrak{R}\,\zeta \geq 1$, wobei $A$ nach $\zeta = 1$, $A_1$ nach $\zeta = 1 + i$, $A_{-1}$ nach $1 - i$ usw. kommt. $\mathfrak{W}_2$ wird durch die Ortsuniformisierenden $\sqrt{w + \frac{1}{2}}$, $\sqrt[\lambda]{w}$ und Anwendung quasikonformer Abbildungen in $\mathfrak{R}\,\zeta^* \geq 0$, $|\mathfrak{J}\,\zeta^*| \leq 1$ schlicht abgebildet. Die Verwendung quasikonformer Abbildungen erlaubt es, die Randabbildung in starkem Maße vorzuschreiben. Die Randabbildung läßt sich hier so einrichten, daß $\zeta = 1 + i\,\varrho$ und $\zeta^* = \varrho - i$ für $\varrho \geq 1$ in denselben Flächenpunkt übergehen und gleiches auch für $\zeta = 1 - i\,\varrho$ und $\zeta^* = \varrho + i$ gilt. Durch einen Schnitt längs der positiv reellen Achse zerfällt $\mathfrak{R}\,\zeta^* \geq 1$, $|\mathfrak{J}\,\zeta^*| \leq 1$ in zwei Teilgebiete $T'$ und $T''$, $T''$: $0 \leq \mathfrak{J}\,\zeta^* \leq 1$. Legt man nun $T'$ und $T''$ so in die $\zeta$-Ebene, daß $\zeta^* = \varrho - i$ bzw. $\zeta^* = \varrho + i$ in $\zeta = 1 + i\varrho$ bzw. $\zeta = 1 - i\varrho$ übergeht, so erhält man in $G_\zeta$: $\mathfrak{R}\,\zeta \geq 0$, $|\zeta| \geq 1$ ein schlichtes Bild von $\overline{\mathfrak{W}} = \mathfrak{W} - \mathfrak{W}'$, wobei $\mathfrak{W}'$ ein endlich vielblättriges Flächenstück ist. Die Punkte $\zeta = i\,\varrho$ und $\zeta = -i\,\varrho$ sind für $\varrho \geqq 1$ zu identifizieren. Die Funktion $z = z(\zeta)$ bildet $G_\zeta$ in $|z| < \infty$ ab, wobei die $\mathfrak{W}'$ entsprechende Umgebung von $z = 0$ auszuschließen ist. Diese Abbildung ist für $\mathfrak{R}\,\zeta > 1$ konform; für $0 \leq \mathfrak{R}\,\zeta \leq 1$ liegt eine stetig differenzierbare Abbildung vor. Die stetige Differenzierbarkeit der Abbildung ist möglicherweise längs Kurvenstücken durchbrochen, die sich nirgends im Endlichen häufen. Ist $\zeta$ ein Punkt, dessen $w$-Bild einem algebraischen Windungsflächenstück mit der Blattzahl $\lambda = 2(\mu + 1)$ angehört, dann gilt für den Dilatationsquotienten bei $\zeta \leftrightarrow z$ $D_{\zeta/z} \leq C_1\,\mu$, wobei $C_1$ eine von $\mu$ unabhängige Konstante ist. Bei der Wahl $\mu = 1, 2, 3, \ldots$ ist $D_{\zeta/z} \leq C_2 \sqrt{\varrho}$, gültig für alle $\varrho \geq 1$. Durch $Z = \zeta^2$ geht $\mathfrak{R}\,\zeta \geq 0$ in die $Z$-Ebene über. In $Z \leftrightarrow z$ erhält man eine quasikonforme Abbildung einer Umgebung von $Z = \infty$ ($|Z| \geq R_0$) in eine Umgebung von $z = \infty$ ($|z| \geq r_0$), die der Bedingung $J(R) = \iint\limits_{|Z| \geq R} (D_{Z/z} - 1)\,\boxed{d \log Z} \to 0$ bei $R \to \infty$ genügt. Wegen

$$\boxed{d \log Z} = 4\,\boxed{d \log \zeta} = 4\,\frac{d\varrho}{\varrho}\,d\Theta \quad \text{und} \quad D_{Z/z} \leqq D_{Z/\zeta} \cdot D_{\zeta/z} = D_{\zeta/z}$$

gilt nämlich

$$J(R) \leqq 4 \int\limits_{\sqrt{R}}^{\infty} \int\limits_{-\pi/2}^{\pi/2} (D_{\zeta/z} - 1)\,\frac{d\varrho}{\varrho}\,d\Theta.$$

Ist $\overline{\Theta}(\varrho) = \frac{\pi}{2} - \Theta(\varrho)$ der zu dem in $0 \leq \Re\,\zeta \leq 1$ gelegenen Bogen von $|\zeta| = \varrho$ gehörige Zentriwinkel, so erhält man wegen $D_{\zeta/s} - 1 = 0$ in $\Re\,\zeta > 1$

$$J(R) < 8\,C_2 \int_{\sqrt{R}}^{\infty} \sqrt{\varrho}\;\overline{\Theta}(\varrho)\,\frac{d\varrho}{\varrho} < C_3 \int_{\sqrt{R}}^{\infty} \frac{d\varrho}{\varrho^{3/2}} \quad \left(\text{weil } \overline{\Theta}(\varrho) \leq \frac{k}{\varrho} \text{ für alle } \varrho \geq \varrho_0\right)$$

oder

$$J(R) < \frac{C_4}{R^{1/4}}, \quad \text{also} \quad \lim_{R \to \infty} J(R) = 0\,.$$

Das bedeutet für die Abbildung $Z \leftrightarrow z$ die Gültigkeit der Verzerrungsformel[9])

$$|z| = a\,|Z|\,(1 + \varepsilon(|Z|)), \quad 0 < a < \infty$$

oder

$$R = A\,r\,(1 + \varepsilon(r)) \quad \text{bzw.} \quad \varrho = \sqrt{A}\,\sqrt{r}\,(1 + \varepsilon(r))\,.$$

Danach lassen sich nun die Anzahlfunktionen $n(r, a), \ldots$ berechnen. Für die Anzahlfunktionen, bezogen auf $\Re\,\zeta \geq 0$, $|\zeta| \leq \varrho$, findet man:

$$\nu(\varrho) = \varrho\,(1 + \varepsilon(\varrho)),\ \nu(\varrho, 0) = \varrho\,(1 + \varepsilon(\varrho)),\ \nu\left(\varrho, -\frac{1}{2}\right) = \varrho\,(1 + \varepsilon(\varrho)),\ \nu(\varrho, \infty) = 0,$$

$$\nu_1(\varrho, a) = \varrho\,(1 + \varepsilon(\varrho)),\ \nu_1\left(\varrho, -\frac{1}{2}\right) = \sqrt{\varrho}\,(1 + \varepsilon(\varrho))$$

und daraus weiter:

$$n(r) = \sqrt{r}\;\alpha(r),\ n(r, 0) = \sqrt{r}\;\alpha(r),\ n\left(r, -\frac{1}{2}\right) = \sqrt{r}\;\alpha(r),\ n(r, \infty) = 0,$$

$$n_1(r, 0) = \sqrt{r}\;\alpha(r),\ n_1\left(r, -\frac{1}{2}\right) = \sqrt[4]{r}\;\alpha(r),\ \lim_{r \to \infty} \alpha(r) = \sqrt{A}\,.$$

Für die durch den Streckenkomplex bestimmte ganze transzendente Funktion erhält man also die folgenden Wertverteilungsgrößen[10]):

$$\delta(\infty) = 1,\ \delta(0) = \delta\left(-\frac{1}{2}\right) = 0$$

$$\vartheta(0) = 1,\ \vartheta(\infty) = \vartheta\left(-\frac{1}{2}\right) = 0$$

und die Ordnung $s = \frac{1}{2}$.

Gilt für die Eckenzahl $\lambda_\mu$ der algebraischen Elementargebiete die Formel $\lambda_\mu = 4\,\mu\,d + 4$, $\mu = 0, 1, 2, \ldots$ und $d$ eine natürliche Zahl, so erhält man $\mu = 0\,(\sqrt{\varrho})$, also $D_{\zeta/s} \leq C\sqrt{\varrho}$. Bei dieser Wahl zeigt, wie leicht zu sehen ist, die durch den Streckenkomplex definierte gebrochene Funktion $w = w(z)$ das am Beispiel $d = 1$ festgestellte Wertverteilungsverhalten:

$$\delta(\infty) = 1,\ \vartheta(0) = 1 \quad \text{und} \quad s = \frac{1}{2}\ ^{11)}.$$

---

[9]) l. c. [2b]) und [2d]).

[10]) Vgl. auch E. Ullrich [1b]).

[11]) Der Übergang von den Grundpunkten $w = 0, -\frac{1}{2}, \infty$ zu beliebigen Grundpunkten $w = a, b, c$ ist wegen der Invarianzeigenschaft von $T(r, w)$ bei linearen Transformationen $w \to L(w)$ erlaubt.

Setzt man dagegen $\lambda_\mu = 2^{\mu+2}$, $\mu = 0, 1, 2, \ldots$, so findet man $D_{\zeta/z} \leq c\,\varrho$. Die durch den Streckenkomplex definierte Fläche ist vom parabolischen Typus, so daß in der Erzeugenden dieser Fläche wieder eine gebrochene Funktion vorliegt. Wegen $D_{\zeta/z} \leq c\,\varrho$ kann man aber nicht mehr wie eben auf die Gültigkeit der Verzerrungsformel schließen und kann damit auch nicht die Anzahlfunktionen auf dem angegebenen Wege bestimmen.

E. Ullrich zeigte unter Ausnützung der beiden Hauptsätze der Wertverteilungslehre, wie man bei hinreichend langsamem Anwachsen der monotonen Folge $\lambda_\mu$ auf $\vartheta(0) = 1$ schließen kann[12]). Die Bemerkungen dieses Abschnittes zeigen, daß die Forderung „hinreichend langsam" sicher erfüllt ist, wenn die Folge $\frac{\lambda_\mu}{2}$ der Blattzahlen von der Form $2\,\mu\,d + 2$ ist.

Nach dem Verhalten der schlichten differentialgeometrischen Abbildungen von $\mathfrak{W}_2$ in ein Gebiet der $\zeta$-Ebene scheint es nicht unwahrscheinlich zu sein, daß bei stark anwachsendem $\lambda_\mu$ eine Wachstumserhöhung eintreten kann.

---

[12]) l. c. [1b]) besonders § 8 Höchstwert des Index.

*(Eingegangen am 25. Mai 1949.)*

Math. Annalen, Bd. 122, S. 47—65 (1950).

# Schlußweisen-Kalküle der Prädikatenlogik.

Von

KURT SCHÜTTE in Göttingen.



## Einleitung.

G. GENTZEN hat in seinen „*Untersuchungen über das logische Schließen*“[1]) festgestellt, daß die logischen Sätze in gewissem Sinne umweglos hergeleitet werden können. Er hat zu diesem Zweck einen Kalkül mit *Sequenzen* entwickelt, in dem die von ihm als „*Schnitte*“ bezeichneten Schlußfiguren eliminiert werden können. Der Sequenzenkalkül zeichnet sich durch eine klare und konsequente Systematik aus, ist aber von einer höheren Struktur als die sonst üblichen logischen Kalküle. Es findet nämlich über der Formalisierung der Logik mittels der logischen Grundsymbole noch eine weitere Formalisierung durch Sequenzen statt, indem die logischen Formeln durch Kommata und durch das Sequenzenzeichen zu höheren Figuren zusammengesetzt werden.

Es soll hier die Gentzensche Entdeckung von der Eliminierbarkeit der Schnitte auch für die einfache Formalisierung der Logik nutzbar gemacht werden. Die im Sequenzenkalkül herstellbare Normalform einer Herleitung, in welcher die einmal eingeführten logischen Zeichen nicht wieder verschwinden, ist allerdings ohne Sequenzen nicht zu erreichen. Zu den logischen Verknüpfungsschlüssen treten ja im Sequenzenkalkül noch Strukturschlüsse. Will man ohne Sequenzen auskommen, so muß man diese Strukturschlüsse durch gewisse umformende Schlüsse ersetzen. Durch diese können aber in einer Herleitung vorher eingeführte logische Zeichen nachträglich wieder fortfallen. Die Logik läßt sich aber auch ohne Sequenzen auf der Grundlage von zwei Schlußweisen entwickeln, welche den Strukturschlüssen und den logischen Verknüpfungsschlüssen entsprechen. Es sind dies:

1. *umformende Schlüsse*, bei denen Ober- und Unterformeln logisch äquivalent[2]) sind,

[1]) Math. Z. 39 (1934).

[2]) D. h. Ober- und Unterformeln sollen nicht nur deduktionsgleich sein, indem jede aus der anderen hergeleitet werden kann, sondern es soll auch die logische Formel gültig sein, welche die Äquivalenz der Ober- und Unterformel ausdrückt.

2. *aufbauende Schlüsse*, bei denen diese Äquivalenz nicht besteht und die Oberformeln echte Teile der Unterformeln sind.

Hierzu müssen die Schnitte, welche ja nicht unter diese beiden Schlußweisen fallen, eliminiert werden. Kalküle, in denen jede Herleitung allein mit umformenden und aufbauenden Schlüssen geführt werden kann, möchte ich als „*genetisch*" bezeichnen. Man gewinnt sie, indem man die logischen Grundformeln auf ein Minimum möglichst einfacher Formeln beschränkt, die logischen Gesetzmäßigkeiten also möglichst nicht in Grundformeln festlegt, sondern hauptsächlich durch geeignete Schlußschemata zur Geltung bringt.

Für die *vollständige Prädikatenlogik* werden im folgenden zwei genetische Kalküle $K_1$, $K_2$ angegeben. Im Kalkül $K_1$ ist ähnlich wie im Sequenzenkalkül das Hauptgewicht auf die aufbauenden Schlüsse gelegt, während die umformenden Schlüsse möglichst einfach gehalten sind. In diesem Kalkül ist die Eliminierbarkeit der Schnitte am leichtesten zu beweisen. Der Kalkül $K_2$ enthält dagegen möglichst starke umformende Schlüsse in Form von „*Umsetzungen*", wodurch die aufbauenden Schlüsse weitgehendst eingeschränkt werden können. Die Herleitungen erhalten dadurch eine einfache Normalform, und die Herleitbarkeit einer Formel kann im wesentlichen mittels Umsetzungen untersucht werden.

Erweiterungen der Kalküle $K_1$, $K_2$ durch *Axiome*, welche keine Quantoren enthalten, können durch geeignete Fassung des Axiomensystems stets so gestaltet werden, daß die Schnitte eliminierbar bleiben, woraus ohne weiteres die *Widerspruchsfreiheit* folgt.

$K_3$ ist ein in der Art von $K_1$ entwickelter Kalkül der *intuitionistischen Prädikatenlogik*. Nach den Grundsätzen von $K_2$ läßt sich ebenfalls ein intuitionistischer Kalkül aufbauen. Er besitzt aber nicht die einfachen Schlußweisen und Normalherleitungen wie $K_2$, weshalb von dessen Darstellung hier abgesehen ist.

## § 1. Der aufbauende Kalkül $K_1$.

Die *Grundzeichen* des Kalküls $K_1$, der die vollständige Prädikatenlogik genetisch darstellt, sind

1. kleine lateinische Buchstaben $a, b, c, \ldots$ für *freie Dingvariablen*, $x, y, z, \ldots$ für *gebundene Dingvariablen*,

2. große lateinische Buchstaben $A, B, C, \ldots$ für *Formelvariablen* (mit und ohne Argumentstellen),

3. die logischen Symbole $\vee$, $^{-}$, $(E.)$.

Außerdem können auch Zeichen für bestimmte Dinge, Funktionen, Prädikate und variable Funktionen auftreten.

„*Terme*" werden rekursiv erklärt als

1. die freien Dingvariablen und Zeichen für bestimmte Dinge,

2. die Zeichen für bestimmte und variable Funktionen, deren Argumentstellen durch Terme ausgefüllt sind.

„*Primformeln*" sind die Formelvariablen und Zeichen für bestimmte Prädikate, deren Argumentstellen durch Terme ausgefüllt sind.

Der Begriff „*Formel*" wird folgendermaßen definiert:

1. Jede Primformel ist eine Formel.

2. Sind $\mathfrak{A}$, $\mathfrak{B}$ Formeln, so ist auch $\mathfrak{A} \vee \mathfrak{B}$ eine Formel (die *Disjunktion* „$\mathfrak{A}$ oder $\mathfrak{B}$").

3. Ist $\mathfrak{A}$ eine Formel, so auch $\overline{\mathfrak{A}}$ (die *Negation* „nicht $\mathfrak{A}$").

4. Ist $\mathfrak{A}(a)$ eine Formel, welche die gebundene Dingvariable $x$ nicht enthält, so ist auch $(E\,x)\,\mathfrak{A}(x)$ eine Formel (die *Existenzformel* „Es gibt ein $x$ mit $\mathfrak{A}(x)$").

Als Mitteilungszeichen für Formeln dienen große deutsche Buchstaben, als Mitteilungszeichen für Terme kleine deutsche Buchstaben.

Mittels „*Grundformeln*" und „*Schlußschematen*" werden „*Herleitungen*" aufgebaut. Das sind stammbaumförmige Figuren von Formeln in folgender Anordnung:

1. An den obersten Stellen der Figur stehen nur Grundformeln.
2. Der Zusammenhang zwischen übereinanderstehenden Formeln ist der eines Schlußschemas.
3. An der untersten Stelle steht die Formel, welche durch die Figur als *hergeleitet* gilt.

*Grundformeln* von $K_1$ sind nur

$$\mathfrak{P} \vee \overline{\mathfrak{P}}$$

mit Primformeln $\mathfrak{P}$.

Die *Schlußschemata* gliedern sich in

*I. Umformende Schlußschemata:*

a) $\dfrac{\mathfrak{M} \vee \mathfrak{A} \vee \mathfrak{B} \vee \mathfrak{N}}{\mathfrak{M} \vee \mathfrak{B} \vee \mathfrak{A} \vee \mathfrak{N}}$ (Vertauschungen) b) $\dfrac{\mathfrak{A} \vee \mathfrak{A} \vee \mathfrak{N}}{\mathfrak{A} \vee \mathfrak{N}}$ (Zusammenziehungen)

*II. Aufbauende Schlußschemata:*

a) $\dfrac{\mathfrak{N}}{\mathfrak{A} \vee \mathfrak{N}}$ (Abschwächungen) b) $\dfrac{\overline{\mathfrak{A}} \vee \mathfrak{N} \quad \overline{\mathfrak{B}} \vee \mathfrak{N}}{\overline{\mathfrak{A} \vee \mathfrak{B}} \vee \mathfrak{N}}$ (Kompositionsschlüsse) c) $\dfrac{\mathfrak{A} \vee \mathfrak{N}}{\overline{\overline{\mathfrak{A}}} \vee \mathfrak{N}}$ (Negationsschlüsse)

d) $\dfrac{\mathfrak{A}(\mathfrak{t}) \vee \mathfrak{N}}{(E\,x)\,\mathfrak{A}(x) \vee \mathfrak{N}}$ e) $\dfrac{\overline{\mathfrak{A}}(a) \vee \mathfrak{N}}{\overline{(E\,x)\,\mathfrak{A}(x)} \vee \mathfrak{N}}$ (mit Variablenbedingung)

(Quantorenschlüsse)

*III. Schnittschema:*

$$\frac{\mathfrak{M} \vee \mathfrak{A} \quad \overline{\mathfrak{A}} \vee \mathfrak{N}}{\mathfrak{M} \vee \mathfrak{N}}.$$

Mehrfache Disjunktionen können wir uns zunächst als nach rechts geklammert vorstellen, so daß z. B. $\mathfrak{M} \vee \mathfrak{A} \vee \mathfrak{B} \vee \mathfrak{N}$ die Formel $\mathfrak{M} \vee (\mathfrak{A} \vee (\mathfrak{B} \vee \mathfrak{N}))$ bedeutet. Auf Grund der Vertauschungsschlüsse besteht aber Assoziativität, so daß die Klammern überflüssig werden.

Die mit $\mathfrak{M}$, $\mathfrak{N}$ bezeichneten Formeln sind die „*Nebenglieder*", die übrigen die „*Hauptglieder*" der betreffenden Schlüsse. Die Nebenglieder können teilweise auch fehlen[3]). Das Hauptglied $\mathfrak{A}$ eines Schnittes ist das „*Schnittglied*".

[3]) Nur die Abschwächungen müssen Nebenglieder enthalten, weil sonst keine Oberformeln vorhanden wären. Im übrigen sollen alle Schlußschemata auch die Fälle mit fehlenden Nebengliedern umfassen. Würden bei einem Schnitt keine Nebenglieder auftreten, so wäre die Unterformel leer. Dieser Fall scheidet aber auf Grund der Widerspruchsfreiheit des Kalküls aus.

Die *Variablenbedingung* für IIe besagt, daß die freie Dingvariable $a$ in der Unterformel nicht vorkommen darf.

Die *Vollständigkeit* des Kalküls $K_1$ ist leicht festzustellen. Zunächst kann mittels der aufbauenden Schlußschemata aus Grundformeln $\mathfrak{P} \vee \overline{\mathfrak{P}}$ successive jede Formel $\mathfrak{A} \vee \overline{\mathfrak{A}}$ hergeleitet werden. Ferner können die weiteren logischen Verknüpfungen explizit definiert werden durch

$$\mathfrak{A} \to \mathfrak{B} \equiv \overline{\mathfrak{A}} \vee \mathfrak{B},$$
$$\mathfrak{A} \,\&\, \mathfrak{B} \equiv \overline{\overline{\mathfrak{A}} \vee \overline{\mathfrak{B}}},$$
$$(x)\,\mathfrak{A}(x) \equiv \overline{(E\,x)}\,\overline{\mathfrak{A}}(x).$$

Hiernach sind alle Grundformeln, die sonst für die Prädikatenlogik angegeben werden, leicht herleitbar.

Das Schema IId stellt eine Quantoren-Abschwächung dar. Es kann ebenso wie die Schemata IIa und III nicht umgekehrt werden. *Umkehrbar* sind dagegen außer den umformenden Schlußschematen auch die aufbauenden Schlußschemata IIb, c, e.

Liegt nämlich eine Herleitung vor für eine Formel

$$\overline{\overline{\mathfrak{A} \vee \mathfrak{B}}} \vee \mathfrak{R},$$

so können wir das Formelglied $\overline{\overline{\mathfrak{A} \vee \mathfrak{B}}}$ in der Herleitung nach oben verfolgen und den stammbaumförmig angeordneten „*Formelbund*"[4]) der Glieder $\overline{\overline{\mathfrak{A} \vee \mathfrak{B}}}$ aufsuchen, welche in der Herleitung diesem Glied der Endformel „*zugehörig*" sind. Dazu rechnen wir mit den betreffenden Gliedern der Unterformel eines Schlusses immer auch die entsprechenden Glieder der Oberformeln. Gabelungen treten ein bei Nebengliedern von Kompositionsschlüssen und bei Hauptgliedern von Zusammenziehungen. Man ersieht aus der Betrachtung der Schlußschemata, daß die Glieder des hier betrachteten Formelbundes aufbauenden Schlüssen und Schnitten nur in Nebengliedern angehören. Oberste Glieder des Formelbundes sind entweder Hauptglieder in den Unterformeln von Kompositionsschlüssen oder Bestandteile der Hauptglieder von Abschwächungen. In Grundformeln können sie ja nicht auftreten, da diese keine negierten Negationen enthalten. Ersetzen wir nun alle Glieder des Formelbundes durch $\overline{\overline{\mathfrak{A}}}$, so bleibt der Herleitungszusammenhang gewahrt, wobei die den Formelbund nach oben abschließenden Kompositionsschlüsse mit den rechten Oberformeln samt deren Herleitungen fortfallen. In dieser Weise gewinnen wir aus der Herleitung von $\overline{\overline{\mathfrak{A} \vee \mathfrak{B}}} \vee \mathfrak{R}$ eine Herleitung von $\overline{\overline{\mathfrak{A}}} \vee \mathfrak{R}$, und zwar allein mittels Ersetzungen von $\overline{\overline{\mathfrak{A} \vee \mathfrak{B}}}$ durch $\overline{\overline{\mathfrak{A}}}$ und mittels Streichungen von Herleitungsteilen. Ebenso ist eine Herleitung von $\overline{\overline{\mathfrak{B}}} \vee \mathfrak{R}$ zu erhalten.

Entsprechend kann aus einer Herleitung von

$$\overline{\overline{\mathfrak{A}}} \vee \mathfrak{R} \text{ bzw. } \overline{(E\,x)}\,\mathfrak{A}(x) \vee \mathfrak{R}$$

eine Herleitung von

$$\mathfrak{A} \vee \mathfrak{R} \text{ bzw. } \overline{\mathfrak{A}}(a) \vee \mathfrak{R}$$

---

[4]) Den entsprechenden Begriff hat GENTZEN verwendet in seiner Arbeit: Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie. Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften, Heft 4, 1938.

gebildet werden, weil ja auch die Glieder $\overline{\mathfrak{A}}$ und $\overline{(E\,x)}\,\mathfrak{A}\,(x)$ nicht in Grundformeln auftreten. Hierbei sind auch nur Ersetzungen von $\overline{\mathfrak{A}}$ bzw. $\overline{(E\,x)}\,\mathfrak{A}\,(x)$ durch $\mathfrak{A}$ bzw. $\overline{\mathfrak{A}}\,(a)$ sowie Streichungen von Schlüssen IIc bzw. IIe vorzunehmen. Bei den Ersetzungen von $\overline{(E\,x)}\,\mathfrak{A}\,(x)$ durch $\overline{\mathfrak{A}}\,(a)$ ist darauf zu achten, daß die IIe-Schlüsse, welche hierdurch beseitigt werden, in den Oberformeln sämtlich das Hauptglied $\overline{\mathfrak{A}}\,(a)$ mit der gleichen freien Dingvariablen $a$ besitzen. Das kann aber auf Grund der hier bestehenden Variablenbedingung immer durch Umbenennungen der Variablen in den betreffenden Herleitungsteilen erreicht werden, ohne daß sich dadurch an der Herleitung sonst etwas ändert.

Die Schlüsse IIb, c, e sind also in der Weise umkehrbar, daß aus Herleitungen von Formeln, welche die Gestalt der betreffenden Unterformeln besitzen, durch gewisse Kürzungen Herleitungen für die entsprechenden Oberformeln zu gewinnen sind.

## § 2. Elimination der Schnitte in $K_1$.

Um den genetischen Charakter des Kalküls $K_1$ (d. h. die Entbehrlichkeit der Schnitte) zu beweisen, nehmen wir an, es sei eine Herleitung gegeben, die einen obersten Schnitt

$$(\Sigma)\quad \frac{\mathfrak{R} \vee \mathfrak{S} \qquad \overline{\mathfrak{S}} \vee \mathfrak{T}}{\mathfrak{R} \vee \mathfrak{T}}$$

(über dem in der Herleitung keine weiteren Schnitte stehen) enthält. Die Herleitung wird dann in Fallunterscheidungen schrittweise so umgeformt, daß der Schnitt $(\Sigma)$ in der Herleitung nach oben rückt oder verschwindet bzw. durch einfachere Schnitte ersetzt wird.

*I. Fall.* Das Schnittglied $\mathfrak{S}$ sei eine Primformel.

Ia) In dem Herleitungsteil oberhalb der rechten Oberformel möge kein aufbauender Schluß auftreten.

Dann ist $\mathfrak{T} \vee \mathfrak{S}$ eine Grundformel, da sich an Grundformeln außer aufbauenden Schlüssen und Schnitten nur eine Vertauschung anschließen kann. Folglich stimmen $\mathfrak{S}$ und $\mathfrak{T}$ überein. Der Schnitt $(\Sigma)$ kann nun mit dem Herleitungsteil der rechten Oberformel gestrichen werden, weil die Unterformel mit der linken Oberformel übereinstimmt.

Ib) Oberhalb der rechten Oberformel möge mindestens ein aufbauender Schluß auftreten.

Die Glieder des Formelbundes von $\overline{\mathfrak{S}}$ können (da $\mathfrak{S}$ prim ist) den aufbauenden Schlüssen nur in Nebengliedern und in Hauptgliedern von Abschwächungen angehören.

Ib 1) Ist der letzte aufbauende Schluß oberhalb $\overline{\mathfrak{S}} \vee \mathfrak{T}$ eine Abschwächung, so hat der Herleitungsteil die Form

$$\begin{array}{ll} \dfrac{\mathfrak{R}}{\mathfrak{A} \vee \mathfrak{R}} & \text{(Abschwächung)} \\ \overline{\mathfrak{S}} \vee \mathfrak{T} & \text{(umformende Schlüsse).} \end{array}$$

Wir streichen hier $\mathfrak{A}$ und alle darunterstehenden entsprechenden Glieder, wobei unter Umständen Zusammenziehungen fortfallen oder verkürzte

Hauptglieder erhalten. Dadurch geht $\overline{\mathfrak{S}} \vee \mathfrak{T}$ entweder in $\overline{\mathfrak{S}} \vee \mathfrak{T}^*$ oder in $\mathfrak{T}^*$ über, wo $\mathfrak{T}^*$ eine (echte oder unechte) Teildisjunktion von $\mathfrak{T}$ ist[5]).

Im ersten Fall ersetzen wir den Herleitungsteil mit dem Schnitt durch

$$\begin{array}{c} \qquad\qquad \mathfrak{N} \\ \mathfrak{R} \vee \mathfrak{S} \qquad \overline{\mathfrak{S}} \vee \mathfrak{T}^* \\ \hline \mathfrak{R} \vee \mathfrak{T}^* \\ \hline \mathfrak{R} \vee \mathfrak{T} \end{array} \quad \begin{array}{l} \text{(umformende Schlüsse)} \\ \text{(Schnitt)} \\ \text{(evtl. Abschwächung mit Vertauschungen)}, \end{array}$$

im zweiten Fall durch

$$\begin{array}{c} \mathfrak{N} \\ \mathfrak{T}^* \\ \hline \mathfrak{R} \vee \mathfrak{T} \end{array} \quad \begin{array}{l} \text{(umformende Schlüsse)} \\ \text{(Abschwächung evtl. mit Vertauschungen)}. \end{array}$$

Hierdurch rückt der Schnitt entweder über die Abschwächung nach oben, oder er fällt mit seinem linken Herleitungsteil ganz fort.

I b 2) Ist der letzte aufbauende Schluß oberhalb $\overline{\mathfrak{S}} \vee \mathfrak{T}$ keine Abschwächung, so hat der Herleitungsteil (sofern der aufbauende Schluß nur eine Oberformel besitzt) die Form

$$\begin{array}{c} \mathfrak{A}_1 \vee \mathfrak{N} \\ \hline \mathfrak{A}_2 \vee \mathfrak{N} \\ \overline{\mathfrak{S}} \vee \mathfrak{T} \end{array} \quad \begin{array}{l} \text{(aufbauender Schluß mit Hauptgliedern } \mathfrak{A}_1, \mathfrak{A}_2\text{)} \\ \text{(umformende Schlüsse)}, \end{array}$$

wobei die zu $\overline{\mathfrak{S}}$ zugehörigen Glieder nur in $\mathfrak{N}$ auftreten. Durch Zusammenziehungen und Vertauschungen können wir aus $\mathfrak{A}_1 \vee \mathfrak{N}$ eine Formel $\overline{\mathfrak{S}} \vee \mathfrak{A}_1 \vee \mathfrak{M}$ gewinnen, in der $\mathfrak{M}$ kein Disjunktionsglied $\overline{\mathfrak{S}}$ enthält. Wir ersetzen dann den Herleitungsteil mit dem Schnitt durch

$$\begin{array}{c} \qquad\qquad \mathfrak{A}_1 \vee \mathfrak{N} \\ \mathfrak{R} \vee \mathfrak{S} \qquad \overline{\mathfrak{S}} \vee \mathfrak{A}_1 \vee \mathfrak{M} \\ \hline \mathfrak{R} \vee \mathfrak{A}_1 \vee \mathfrak{M} \\ \hline \mathfrak{R} \vee \mathfrak{A} \vee \mathfrak{M} \\ \hline \mathfrak{R} \vee \mathfrak{T} \end{array} \quad \begin{array}{l} \text{(umformende Schlüsse)} \\ \text{(Schnitt)} \\ \text{(aufbauender Schluß mit Vertauschungen)} \\ \text{(evtl. umformende Schlüsse mit Abschwächung)}. \end{array}$$

Hat der aufbauende Schluß zwei Oberformeln (Kompositionsschluß), so ist der Schnitt entsprechend in der Ersatzfigur doppelt aufzunehmen. In jedem Fall rückt der Schnitt über den aufbauenden Schluß nach oben.

*II. Fall.* Das Schnittglied $\mathfrak{S}$ sei keine Primformel.

II a) Ist $\mathfrak{S}$ von der Form $\mathfrak{A} \vee \mathfrak{B}$, so läßt sich, wie wir gesehen haben, die Herleitung der rechten Oberformel $\overline{\mathfrak{A} \vee \mathfrak{B}} \vee \mathfrak{T}$ umwandeln in Herleitungen von $\overline{\mathfrak{A}} \vee \mathfrak{T}$ und $\overline{\mathfrak{B}} \vee \mathfrak{T}$. Diese sind ebenso wie der ursprüngliche Herleitungsteil schnittfrei. Wir ersetzen den Schnitt

$$\frac{\mathfrak{R} \vee \mathfrak{A} \vee \mathfrak{B} \qquad \overline{\mathfrak{A} \vee \mathfrak{B}} \vee \mathfrak{T}}{\mathfrak{R} \vee \mathfrak{T}} \quad \text{(Schnitt mit Schnittglied } \mathfrak{A} \vee \mathfrak{B}\text{)}$$

durch

---

[5]) $\mathfrak{T}^*$ entsteht aus $\mathfrak{T}$ durch Streichungen einzelner Disjunktionsglieder oder stimmt mit $\mathfrak{T}$ überein.

$$\begin{array}{c}\dfrac{\begin{array}{c}\dfrac{\mathfrak{R}\vee\mathfrak{A}\vee\mathfrak{B}\qquad\overline{\mathfrak{B}}\vee\mathfrak{T}}{\mathfrak{R}\vee\mathfrak{A}\vee\mathfrak{T}}\\ \mathfrak{R}\vee\mathfrak{T}\vee\mathfrak{A}\end{array}\qquad\overline{\mathfrak{A}}\vee\mathfrak{T}}{\mathfrak{R}\vee\mathfrak{T}\vee\mathfrak{T}}\\ \mathfrak{R}\vee\mathfrak{T}\end{array}\qquad\begin{array}{l}\text{(Schnitt mit Schnittglied }\mathfrak{B})\\ \text{(Vertauschungen)}\\ \text{(Schnitt mit Schnittglied }\mathfrak{A})\\ \text{(umformende Schlüsse)}\end{array}$$

IIb) Ist $\mathfrak{S}$ von der Form $\overline{\mathfrak{A}}$, so bilden wir aus der Herleitung von $\overline{\overline{\mathfrak{A}}}\vee\mathfrak{T}$ eine Herleitung von $\mathfrak{A}\vee\mathfrak{T}$ und ersetzen den Schnitt

$$\dfrac{\mathfrak{R}\vee\overline{\mathfrak{A}}\qquad\overline{\overline{\mathfrak{A}}}\vee\mathfrak{T}}{\mathfrak{R}\vee\mathfrak{T}}\qquad\text{(Schnitt mit Schnittglied }\overline{\mathfrak{A}})$$

durch

$$\begin{array}{c}\dfrac{\begin{array}{cc}\mathfrak{A}\vee\mathfrak{T} & \mathfrak{R}\vee\overline{\mathfrak{A}}\\ \mathfrak{T}\vee\mathfrak{A} & \overline{\mathfrak{A}}\vee\mathfrak{R}\end{array}}{\mathfrak{T}\vee\mathfrak{R}}\\ \mathfrak{R}\vee\mathfrak{T}\end{array}\qquad\begin{array}{l}\text{(Vertauschungen)}\\ \text{(Schnitt mit Schnittglied }\mathfrak{A})\\ \text{(Vertauschung)}\end{array}$$

IIc) Ist schließlich $\mathfrak{S}$ von der Form $(E\,x)\,\mathfrak{A}(x)$, so können wir aus der Herleitung von $\overline{(E\,x)}\,\mathfrak{A}(x)\vee\mathfrak{T}$ eine Herleitung von $\overline{\mathfrak{A}}(a)\vee\mathfrak{T}$[6]) und hieraus Herleitungen von $\overline{\mathfrak{A}}(\mathfrak{t})\vee\mathfrak{T}$ für beliebige Terme $\mathfrak{t}$ bilden, indem wir $a$ überall durch $\mathfrak{t}$ ersetzen. Wir betrachten dann den Formelbund der zu $(E\,x)\,\mathfrak{A}(x)$ zugehörigen Glieder oberhalb der linken Oberformel $\mathfrak{R}\vee(E\,x)\,\mathfrak{A}(x)$. Die obersten Glieder desselben stehen entweder als Hauptglieder von IId-Schlüssen in Unterformeln

$$(E\,x)\,\mathfrak{A}(x)\vee\mathfrak{R}_i$$

mit den Oberformeln

$$\mathfrak{A}(\mathfrak{t}_i)\vee\mathfrak{R}_i,$$

oder in Hauptgliedern von Abschwächungen; während alle übrigen Glieder des Formelbundes den aufbauenden Schlüssen nur in Nebengliedern angehören. Wir ersetzen einen jeden solchen IId-Schluß durch

$$\begin{array}{c}\dfrac{\begin{array}{c}\mathfrak{A}(\mathfrak{t}_i)\vee\mathfrak{R}_i\\ \mathfrak{R}_i\vee\mathfrak{A}(\mathfrak{t}_i)\end{array}\qquad\overline{\mathfrak{A}}(\mathfrak{t}_i)\vee\mathfrak{T}}{\mathfrak{R}_i\vee\mathfrak{T}}\\ \mathfrak{T}\vee\mathfrak{R}_i\end{array}\qquad\begin{array}{l}\text{(Vertauschung)}\\ \text{(Schnitt mit Schnittglied }\mathfrak{A}(\mathfrak{t}_i))\\ \text{(Vertauschung).}\end{array}$$

Werden weiterhin alle Glieder des Formelbundes durch $\mathfrak{T}$ ersetzt, so bleibt der Herleitungszusammenhang gewahrt, und die linke Oberformel des Schnittes $(\Sigma)$ geht über in $\mathfrak{R}\vee\mathfrak{T}$, wodurch dieser Schnitt überflüssig wird.

In den Fällen IIa) bis IIc) ist der Schnitt $(\Sigma)$ beseitigt. Die dabei neu hinzugetretenen Schnitte haben ausnahmslos Schnittglieder, welche weniger logische Zeichen als das ursprüngliche Schnittglied $\mathfrak{S}$ enthalten. In den Fällen I ist dagegen der Schnitt $(\Sigma)$ entweder beseitigt oder über aufbauende Schlüsse nach oben hinweggezogen worden, ohne daß neue Schnitte hinzugetreten sind.

Man erkennt, daß das angegebene Ersetzungsverfahren immer eine gewisse Reduktion der Schnitte herbeiführt, welche nach fortgesetzter Wiederholung schließlich zur Beseitigung aller Schnitte führt. Folglich ist jede unter Verwendung von Schnitten hergeleitete Formel auch ohne Schnitte herleitbar.

---

[6]) Dabei soll die freie Variable $a$ so gewählt sein, daß sie in $\overline{(E\,x)}\,\mathfrak{A}(x)\vee\mathfrak{T}$ nicht vorkommt.

## § 3. Erweiterungen von $K_1$ durch Axiome ohne Quantoren.

Bei der Eliminierbarkeit der Schnitte andelt es sich nicht um eine Herleitbarkeit des Schnittschemas aus den übrigen Schlußschematen und Grundformeln des Kalküls. Die Eliminierbarkeit ist vielmehr wesentlich an den betreffenden Kalkül, für den sie nachgewiesen ist, gebunden und kann bei Erweiterungen des Kalküls hinfällig werden. Solche Erweiterungen finden statt, wenn eine bestimmte Theorie auf prädikatenlogischer Grundlage axiomatisch festgelegt wird. Es treten dann zu den logischen Grundformeln noch gewisse *Axiome*, welche sich auf *bestimmte Prädikate* und *bestimmte Dinge* beziehen, als weitere Grundformeln hinzu. Soll nun der genetische Charakter unseres Kalküls $K_1$ eine über die engere formale Logik hinausgehende Bedeutung haben, so kommt es darauf an, daß er bei der Hinzunahme von Axiomen erhalten bleibt bzw. wieder hergestellt werden kann.

Für *Axiomensysteme*, welche *ohne Quantoren* darstellbar sind, ist dies tatsächlich der Fall. Dazu muß das Axiomensystem nur in einer bestimmten Form gefaßt werden.

Zunächst beschränken wir die Axiome auf Disjunktionen aus Primformeln und einfach negierten Primformeln. Wir wollen sie als „*Primärformeln*" bezeichnen. Die genaue Definition lautet:

1. Jede Primformel ist eine Primärformel.
2. Ist $\mathfrak{P}$ eine Primformel, so ist $\overline{\mathfrak{P}}$ eine Primärformel.
3. Sind $\mathfrak{P}$, $\mathfrak{Q}$ Primärformeln, so ist auch $\mathfrak{P} \vee \mathfrak{Q}$ eine Primärformel.

Durch die Beschränkung auf Primärformeln ist das Axiomensystem nicht eingeengt. Man kann ja unter Hinzunahme des Konjunktionszeichens jede Formel, die axiomatisch eingeführt werden soll, in konjunktiver Normalform entwickeln. Die einzelnen Glieder dieser Normalform sind Primärformeln. Nimmt man diese als Axiome auf, so ist die verlangte Formel herleitbar.

Ferner haben wir noch folgende Forderungen an das Axiomensystem zu stellen:

1. Ist $\mathfrak{A}$ ein Axiom, so auch jede Formel $\mathfrak{A}^*$, welche durch Vertauschungen und Zusammenziehungen aus $\mathfrak{A}$ entsteht.
2. Sind $\mathfrak{M} \vee \mathfrak{A}$ und $\overline{\mathfrak{A}} \vee \mathfrak{N}$ Axiome, so ist auch $\mathfrak{M} \vee \mathfrak{N}$ ein Axiom.

Die 2. Bedingung soll auch die Fälle mit fehlendem $\mathfrak{M}$ oder $\mathfrak{N}$ umfassen. Der Fall, daß $\mathfrak{M}$ und $\mathfrak{N}$ zugleich fehlen, scheidet bei widerspruchsfreien Axiomensystemen aus. Ein System, das den beiden angegebenen Bedingungen genügt, das also gegenüber umformenden Schlüssen und gegenüber Schnitten abgeschlossen ist, wollen wir kurz als „*abgeschlossen*" bezeichnen.

Offenbar läßt sich jedes gegebene Axiomensystem zur Abgeschlossenheit ergänzen, ohne daß hierdurch eine Erweiterung des Formalismus stattfindet. Es werden ja nur Formeln, die ohnehin herleitbar sind, als Axiome aufgenommen. Wir brauchen diese Formeln aber als Axiome, damit die Herleitungen ohne Schnitte durchgeführt werden können. Dabei lassen wir auch Systeme von unendlich vielen Axiomen zu, wenn sie nur in entscheidbarer Weise eindeutig definiert sind. In einem abgeschlossenen Axiomensystem sind natürlich die einzelnen Axiome im allgemeinen nicht voneinander unabhängig.

Erweitern wir nun $K_1$ durch ein abgeschlossenes Axiomensystem von Primärformeln, so kann die Eliminierbarkeit der Schnitte im wesentlichen wie vorher bewiesen werden. Es ist nur folgendes zu ergänzen.

Ist das Schnittglied $\mathfrak{S}$ des betrachteten Schnittes ($\Sigma$) eine Primformel, und treten über einer der beiden Oberformeln aufbauende Schlüsse auf, so vertauschen wir einen letzten aufbauenden Schluß mit dem Schnitt, wie dies in § 2 angegeben ist, und zwar auch dann, wenn nur über der linken Oberformel ein aufbauender Schluß steht. Es kann ja über der linken Oberformel ebenso wie über der rechten verfahren werden.

Tritt dagegen über keiner der beiden Oberformeln ein aufbauender Schluß auf, so ist entweder eine Oberformel (bis auf eine Vertauschung) eine logische Grundformel und der Schnitt wird entbehrlich (da dann die Unterformel mit der anderen Oberformel übereinstimmt), oder beide Oberformeln gehen durch umformende Schlüsse aus Axiomen hervor. Auf Grund der Abgeschlossenheit des Axiomensystems sind dann diese Oberformeln $\mathfrak{R} \vee \mathfrak{S}$, $\overline{\mathfrak{S}} \vee \mathfrak{T}$ selbst Axiome, und es ist auch die Unterformel $\mathfrak{R} \vee \mathfrak{T}$ ein Axiom. Der Schnitt kann also auch in diesem Falle fortfallen.

Im übrigen bleibt das Ersetzungsverfahren des § 2 unverändert bestehen.

Der Kalkül $K_1$ behält also bei Erweiterungen durch abgeschlossene Axiomensysteme von Primärformeln seinen genetischen Charakter.

## § 4. Der Umsetzungs-Kalkül $K_2$.

Nachdem die Eliminierbarkeit der Schnitte für $K_1$ nachgewiesen ist, braucht der entsprechende Beweis für einen anderen logischen Kalkül nicht mehr geführt zu werden, wenn dieser folgende Eigenschaften besitzt:

1. Der Kalkül geht nicht über die Prädikatenlogik hinaus und enthält nur umformende und aufbauende Schlußweisen.

2. In dem Kalkül sind die Grundformeln, umformenden und aufbauenden Schlüsse von $K_1$ enthalten oder herleitbar.

Ein solcher Kalkül ist ja zugleich mit den Erweiterungen, in denen $K_1$ genetisch bleibt, mit $K_1$ äquivalent und braucht in diesen Grenzen das Schnittschema nicht zu enthalten.

Von dieser Art ist der folgende *Kalkül* $K_2$.

Seine *logischen Grundzeichen* sind die Zeichen für die

| | | |
|---|---|---|
| *Disjunktion* | $\mathfrak{A} \vee \mathfrak{B}$ | („$\mathfrak{A}$ oder $\mathfrak{B}$“), |
| *Konjunktion* | $\mathfrak{A} \mathbin{\&} \mathfrak{B}$ | („$\mathfrak{A}$ und $\mathfrak{B}$“), |
| *Negation* | $\overline{\mathfrak{A}}$ | („nicht $\mathfrak{A}$“), |
| *Existenzformel* | $(E\,x)\,\mathfrak{A}(x)$ | („Es gibt ein $x$ mit $\mathfrak{A}(x)$“), |
| *Allformel* | $(x)\,\mathfrak{A}(x)$ | („Für alle $x$ gilt $\mathfrak{A}(x)$“). |

Die sonstige Symbolik ist die von $K_1$. Die Begriffe „*Term*“, „*Primformel*“ und „*Formel*“ sind den hier verwendeten logischen Grundzeichen entsprechend zu verstehen.

*Herleitungen* werden gebildet aus *Grundformeln, Umsetzungsschematen* und *Schlußschematen.* Die Umsetzungsschemata haben die Form

$$\mathfrak{A} = \mathfrak{B}$$

und bedeuten, daß jeder *Ausdruck* der Form $\mathfrak{A}$, welcher als Bestandteil einer Formel auftritt, in der Formel durch den entsprechenden Ausdruck $\mathfrak{B}$ ersetzt werden darf, und daß ebenso $\mathfrak{B}$ durch $\mathfrak{A}$ ersetzbar ist. Hierbei

brauchen $\mathfrak{A}$, $\mathfrak{B}$ keine Formeln zu sein, sondern können auch Ausdrücke sein, welche aus Formeln entstehen, wenn darin freie Dingvariablen durch gebundene ersetzt werden, ohne daß hierzu die bindenden Quantoren (Existenz- oder Allzeichen) innerhalb $\mathfrak{A}$, $\mathfrak{B}$ auftreten. Jede Formel geht durch derartige Umsetzungen wieder in eine Formel über. Die Umsetzungsschemata nehmen hier die Rolle der umformenden Schlußschemata von $K_1$ ein, sind jedoch viel stärker, so daß dadurch die aufbauenden Schlußschemata (hier kurz als „Schlußschemata" bezeichnet) beschränkt werden können.

Die einzelnen Schemata von $K_2$ sind:

*I. Grundformelschema:* $\mathfrak{P} \vee \overline{\mathfrak{P}}$ (mit Primformeln $\mathfrak{P}$).

*II. Umsetzungsschemata:*

| | | |
|---|---|---|
| 1. | $\mathfrak{A} \vee \mathfrak{A} \equiv \mathfrak{A}$ | Idempotenz |
| 2. | $\mathfrak{A} \vee \mathfrak{B} \equiv \mathfrak{B} \vee \mathfrak{A}$ | Kommutativität |
| 3. | $\mathfrak{A} \vee (\mathfrak{B} \vee \mathfrak{C}) \equiv (\mathfrak{A} \vee \mathfrak{B}) \vee \mathfrak{C}$ | Assoziativität |
| 4. | $(\mathfrak{A} \,\&\, \mathfrak{B}) \vee \mathfrak{C} \equiv (\mathfrak{A} \vee \mathfrak{C}) \,\&\, (\mathfrak{B} \vee \mathfrak{C})$ | Distributivität |
| 5. | $\overline{\mathfrak{A} \vee \mathfrak{B}} \equiv \overline{\mathfrak{A}} \,\&\, \overline{\mathfrak{B}}$ | Und-Oder-Umwandlung |
| 6. | $\overline{\overline{\mathfrak{A}}} \equiv \mathfrak{A}$ | Doppelte Negation |
| 7. | $(E\,x)\,[\mathfrak{A}(x) \vee \mathfrak{B}] \equiv (E\,x)\,\mathfrak{A}(x) \vee \mathfrak{B}$ | Quantoren-Assoziativität |
| 8. | $(x)\,[\mathfrak{A}(x) \vee \mathfrak{B}] \equiv (x)\,\mathfrak{A}(x) \vee \mathfrak{B}$ | Quantoren-Distributivität |
| 9. | $\overline{(E\,x)}\,\mathfrak{A}(x) \equiv (x)\,\overline{\mathfrak{A}}(x)$ | Quantoren-Umwandlung. |

*III. Schlußschemata:*

| | | | |
|---|---|---|---|
| a) $\dfrac{\mathfrak{A}}{\mathfrak{A} \vee \mathfrak{B}}$ | b) $\dfrac{\mathfrak{A} \quad \mathfrak{B}}{\mathfrak{A} \,\&\, \mathfrak{B}}$ | c) $\dfrac{\mathfrak{A}(t)}{(E\,x)\,\mathfrak{A}(x)}$ | d) $\dfrac{\mathfrak{A}(a)}{(x)\,\mathfrak{A}(x)}$ |
| Disjunktionsschl. | Konjunktionsschl. | Existenzschl. | Allschluß (mit Variablenbedingung) |

Bei den Umsetzungsschematen 7 und 8 soll $\mathfrak{B}$ die gebundene Variable $x$ nicht enthalten.

Die hier angegebenen (aufbauenden) Schlußschemata sind die einfachsten und wenigsten, die ein genetischer Kalkül überhaupt enthalten kann. Es muß ja an aussagenlogischen Schlußschematen mindestens eine Abschwächung und ein Schema mit zwei Oberformeln vorhanden sein. Ferner müssen die Quantorenschlüsse mindestens in einer abschwächenden Form und in einer Form mit Variablenbedingung auftreten. Die Grundverknüpfungen „Oder", „Und", „Es gibt" und „Alle" geben diese Schlußweisen am angemessensten wieder; und es kommt ja auch im Sprachgebrauch diesen Beziehungen nicht umsonst eine vorherrschende Bedeutung zu. Eine andere Rolle spielt hier die Negation. Wir brauchen sie im Grundformelschema (dem „Satz vom ausgeschlossenen Dritten") und in den Umsetzungsschematen, wo sie die Umwandlungen zwischen „Und" und „Oder" und zwischen „Alle" und „Es gibt" vermittelt.

Mit Hilfe der Umsetzungsschemata 5, 6, 9 sind aus den übrigen Umsetzungsschematen folgende herleitbar:

| | | | |
|---|---|---|---|
| 1*. | $\mathfrak{A} \,\&\, \mathfrak{A} \equiv \mathfrak{A}$ | 5*. | $\overline{\mathfrak{A} \,\&\, \mathfrak{B}} \equiv \overline{\mathfrak{A}} \vee \overline{\mathfrak{B}}$ |
| 2*. | $\mathfrak{A} \,\&\, \mathfrak{B} \equiv \mathfrak{B} \,\&\, \mathfrak{A}$ | 7*. | $(x)\,[\mathfrak{A}(x) \,\&\, \mathfrak{B}] \equiv (x)\,\mathfrak{A}(x) \,\&\, \mathfrak{B}$ |
| 3*. | $\mathfrak{A} \,\&\, (\mathfrak{B} \,\&\, \mathfrak{C}) \equiv (\mathfrak{A} \,\&\, \mathfrak{B}) \,\&\, \mathfrak{C}$ | 8*. | $(E\,x)\,[\mathfrak{A}(x) \,\&\, \mathfrak{B}] \equiv (E\,x)\,\mathfrak{A}(x) \,\&\, \mathfrak{B}$ |
| 4*. | $(\mathfrak{A} \vee \mathfrak{B}) \,\&\, \mathfrak{C} \equiv (\mathfrak{A} \,\&\, \mathfrak{C}) \vee (\mathfrak{B} \,\&\, \mathfrak{C})$ | 9*. | $\overline{(x)}\,\mathfrak{A}(x) \equiv (E\,x)\,\overline{\mathfrak{A}}(x)$. |

Von diesen ist jedes mit dem entsprechenden Umsetzungsschema gleicher Nummer äquivalent und kann auch an dessen Stelle gesetzt werden.

Die Herleitbarkeit von $K_1$ aus $K_2$ ist folgendermaßen einzusehen. Die *Grundformeln* und *Abschwächungen* von $K_1$ treten auch in $K_2$ auf. Die *umformenden Schlüsse* und *Negationsschlüsse* von $K_1$ sind in $K_2$ herleitbar auf Grund der Assoziativität, Kommutativität, Idempotenz und des Schemas der doppelten Negation. Die übrigen aufbauenden Schlußschemata von $K_1$ können in $K_2$ folgendermaßen hergeleitet werden:

IIb) Aus $\overline{\mathfrak{A}} \vee \mathfrak{R}$, $\overline{\mathfrak{B}} \vee \mathfrak{R}$ folgt durch einen Konjunktionsschluß $(\overline{\mathfrak{A}} \vee \mathfrak{R})\,\&$ $\&\,(\overline{\mathfrak{B}} \vee \mathfrak{R})$, auf Grund der Distributivität $(\overline{\mathfrak{A}} \,\&\, \overline{\mathfrak{B}}) \vee \mathfrak{R}$ und durch Und-Oder-Umwandlung $\overline{\mathfrak{A} \vee \mathfrak{B}} \vee \mathfrak{R}$.

IId) Aus $\mathfrak{A}(t) \vee \mathfrak{R}$ folgt durch einen Existenzschluß $(E\,x)\,[\mathfrak{A}(x) \vee \mathfrak{R}]$ und auf Grund der Quantoren-Assoziativität $(E\,x)\,\mathfrak{A}(x) \vee \mathfrak{R}$.

IIe) Aus $\overline{\mathfrak{A}}(a) \vee \mathfrak{R}$ folgt durch einen Allschluß $(x)\,[\overline{\mathfrak{A}}(x) \vee \mathfrak{R}]$, auf Grund der Quantoren-Distributivität $(x)\,\overline{\mathfrak{A}}(x) \vee \mathfrak{R}$ und durch Quantoren-Umwandlung $\overline{(E\,x)}\,\mathfrak{A}(x) \vee \mathfrak{R}$.

Ferner sind die *expliziten Definitionen* von $K_1$

$$\mathfrak{A} \,\&\, \mathfrak{B} \equiv \overline{\overline{\mathfrak{A}} \vee \overline{\mathfrak{B}}}, \qquad (x)\,\mathfrak{A}(x) \equiv \overline{(E\,x)}\,\overline{\mathfrak{A}}(x)$$

als Umsetzungen in $K_2$ vorhanden.

Damit ist die Äquivalenz von $K_2$ mit $K_1$ nachgewiesen.

## § 5. Unabhängigkeit der Umsetzungsschemata.

Die Umsetzungsschemata von $K_2$ sind alle voneinander (und von den Grundformeln und Schlußschematen) *unabhängig*, wie sich mittels der *Methode der Wertung*[7]) nachweisen läßt.

Für die Unabhängigkeitsbeweise der Umsetzungsschemata 1, 5, 6, 9 genügen zweifache Wertungen mit den Werten $\alpha$ („richtig") und $\beta$ („falsch") und zwar folgende:

1) $\mathfrak{A} \vee \mathfrak{B} = \alpha$, $\mathfrak{A} \,\&\, \mathfrak{B} = \alpha$, $\overline{\mathfrak{A}}$ gleichwertig mit $\mathfrak{A}$,

$$(E\,x)\,\mathfrak{A}(x) = \begin{cases} \alpha, & \text{wenn es ein } x \text{ mit } \mathfrak{A}(x) = \alpha \text{ gibt,} \\ \beta & \text{sonst,} \end{cases}$$

$$(x)\,\mathfrak{A}(x) = \begin{cases} \alpha, & \text{wenn } \mathfrak{A}(x) = \alpha \text{ für alle } x, \\ \beta & \text{sonst.} \end{cases}$$

Hierbei ist die *Idempotenz* nicht erfüllt, da $\beta \vee \beta = \alpha$ ist.

5) $$\mathfrak{A} \vee \mathfrak{B} = \begin{cases} \alpha, & \text{wenn } \mathfrak{A} = \alpha \text{ oder } \mathfrak{B} = \alpha, \\ \beta & \text{sonst,} \end{cases}$$

$\mathfrak{A} \,\&\, \mathfrak{B} = \alpha$, $\overline{\mathfrak{A}}$ ungleichwertig mit $\mathfrak{A}$,

Quantorenwertung wie vorher.

Hiernach ist $\overline{\overline{\alpha} \vee \overline{\alpha}} = \beta$, $\overline{\alpha} \,\&\, \overline{\alpha} = \alpha$, was gegen die *Und-Oder-Umwandlung* verstößt.

---

[7]) Vgl. hierzu Hilbert-Bernays, Grundlagen der Mathematik, Bd. I, S. 72—82.

$$6)\quad \mathfrak{A} \vee \mathfrak{B} = \begin{cases} \alpha, \text{ wenn } \mathfrak{A} = \alpha \text{ oder } \mathfrak{B} = \alpha, \\ \beta \text{ sonst,} \end{cases}$$

$$\mathfrak{A} \mathbin{\&} \mathfrak{B} = \begin{cases} \alpha, \text{ wenn } \mathfrak{A} = \alpha \text{ und } \mathfrak{B} = \alpha, \\ \beta \text{ sonst,} \end{cases}$$

$$\overline{\mathfrak{A}} = \alpha, \text{ Quantorenwertung wie vorher.}$$

Das Schema der *doppelten Negation* ist nicht erfüllt; denn es ist $\overline{\overline{\beta}} = \alpha$.

$$9)\quad \mathfrak{A} \vee \mathfrak{B},\ \mathfrak{A} \mathbin{\&} \mathfrak{B} \text{ wie unter 6),}$$

$$\overline{\mathfrak{A}} \text{ ungleichwertig mit } \mathfrak{A},$$

$$(E\,x)\,\mathfrak{A}(x) = \alpha,\ (x)\,\mathfrak{A}(x) = \alpha.$$

Daraus folgt $\overline{(E\,x)}\,\mathfrak{A}(x) = \beta$, $(x)\,\overline{\mathfrak{A}}(x) = \alpha$, was gegen die *Quantoren-Umwandlung* verstößt.

Bei allen diesen Wertungen erhalten die Grundformeln stets den Wert $\alpha$, und es sind mit Hilfe der Schlußschemata nur Formeln vom Wert $\alpha$ herleitbar. Bis auf die jeweils angegebenen Ausnahmen sind die Umsetzungsschemata in der Weise erfüllt, daß auf beiden Seiten immer gleichwertige Formeln stehen.

Um die Unabhängigkeit der Umsetzungsschemata 2, 7, 8 nachzuweisen, sind dreifache Wertungen erforderlich. Wir verwenden zwei ausgezeichnete Werte $\alpha$, $\beta$ und einen weiteren Wert $\gamma$. Die einzelnen Wertungen sind:

2)

$$\begin{array}{cc|ccc} & & & \mathfrak{B} & \\ \mathfrak{A} \vee \mathfrak{B} & & \alpha & \beta & \gamma \\ \hline & \alpha & \alpha & \alpha & \alpha \\ \mathfrak{A} & \beta & \alpha & \beta & \beta \\ & \gamma & \alpha & \gamma & \gamma \end{array} \qquad \begin{array}{cc|ccc} & & & \mathfrak{B} & \\ \mathfrak{A} \mathbin{\&} \mathfrak{B} & & \alpha & \beta & \gamma \\ \hline & \alpha & \alpha & \alpha & \gamma \\ \mathfrak{A} & \beta & \beta & \beta & \gamma \\ & \gamma & \gamma & \gamma & \gamma \end{array} \qquad \begin{array}{c|c} \mathfrak{A} & \overline{\mathfrak{A}} \\ \hline \alpha & \gamma \\ \beta & \beta \\ \gamma & \alpha \end{array}$$

$$(E\,x)\,\mathfrak{A}(x) = (x)\,\mathfrak{A}(x) = \begin{cases} \alpha, \text{ wenn } \mathfrak{A}(x) = \alpha \text{ für alle } x, \\ \gamma, \text{ wenn } \mathfrak{A}(x) = \gamma \text{ für alle } x, \\ \beta \text{ sonst.} \end{cases}$$

Hier ist die *Kommutativität* durchbrochen bei $\beta \vee \gamma = \beta$, $\gamma \vee \beta = \gamma$.

7)

$$\begin{array}{cc|ccc} & & & \mathfrak{B} & \\ \mathfrak{A} \vee \mathfrak{B} & & \alpha & \beta & \gamma \\ \hline & \alpha & \alpha & \alpha & \alpha \\ \mathfrak{A} & \beta & \alpha & \beta & \beta \\ & \gamma & \alpha & \beta & \gamma \end{array} \qquad \begin{array}{cc|ccc} & & & \mathfrak{B} & \\ \mathfrak{A} \mathbin{\&} \mathfrak{B} & & \alpha & \beta & \gamma \\ \hline & \alpha & \alpha & \beta & \gamma \\ \mathfrak{A} & \beta & \beta & \beta & \gamma \\ & \gamma & \gamma & \gamma & \gamma \end{array} \qquad \begin{array}{c|c} \mathfrak{A} & \overline{\mathfrak{A}} \\ \hline \alpha & \gamma \\ \beta & \beta \\ \gamma & \alpha \end{array}$$

$$(E\,x)\,\mathfrak{A}(x) = \begin{cases} \beta, \text{ wenn es ein } x \text{ mit } \mathfrak{A}(x) \neq \gamma \text{ gibt,} \\ \gamma \text{ sonst,} \end{cases}$$

$$(x)\,\mathfrak{A}(x) = \begin{cases} \alpha, \text{ wenn } \mathfrak{A}(x) = \alpha \text{ für alle } x, \\ \beta \text{ sonst.} \end{cases}$$

Daraus folgt $(E\,x)\,[\mathfrak{A}(x) \vee \alpha] = \beta$, $(E\,x)\,\mathfrak{A}(x) \vee \alpha = \alpha$, was gegen die *Quantoren-Assoziativität* verstößt.

8) $\mathfrak{A} \vee \mathfrak{B}$, $\mathfrak{A} \mathbin{\&} \mathfrak{B}$, $\overline{\mathfrak{A}}$ wie unter 7),

$$(E\,x)\,\mathfrak{A}(x) = \begin{cases} \alpha, \text{ wenn es ein } x \text{ mit } \mathfrak{A}(x) = \alpha \text{ gibt,} \\ \beta \text{ sonst,} \end{cases}$$

$$(x)\,\mathfrak{A}(x) = \begin{cases} \beta, \text{ wenn } \mathfrak{A}(x) \neq \gamma \text{ für alle } x, \\ \gamma \text{ sonst.} \end{cases}$$

Die *Quantoren-Distributivität* ist nicht erfüllt wegen $(x)\,[\mathfrak{A}(x) \vee \alpha] = \beta$, $(x)\,\mathfrak{A}(x) \vee \alpha = \alpha$.

Bei den letzten drei Wertungen sind immer nur Formeln mit den ausgezeichneten Werten $\alpha$, $\beta$ herleitbar. Die Umsetzungsschemata sind bis auf die angegebenen Ausnahmen erfüllt.

Die Unabhängigkeitsnachweise für die Assoziativität und Distributivität erfordern vierfache Wertungen. Wir nehmen $\alpha$, $\beta$, $\gamma$ als ausgezeichnete Werte und $\delta$ als nicht ausgezeichneten Wert. Die Wertungen sind:

3)

$$\begin{array}{cc|cccc}
 & \mathfrak{A} \vee \mathfrak{B} & \alpha & \beta & \gamma & \delta \\
\hline
\mathfrak{A} & \alpha & \alpha & \alpha & \beta & \beta \\
 & \beta & \alpha & \beta & \beta & \beta \\
 & \gamma & \beta & \beta & \gamma & \gamma \\
 & \delta & \beta & \beta & \gamma & \delta
\end{array}
\qquad
\begin{array}{cc|cccc}
 & \mathfrak{A} \mathbin{\&} \mathfrak{B} & \alpha & \beta & \gamma & \delta \\
\hline
\mathfrak{A} & \alpha & \alpha & \beta & \gamma & \gamma \\
 & \beta & \beta & \beta & \gamma & \gamma \\
 & \gamma & \gamma & \gamma & \gamma & \delta \\
 & \delta & \gamma & \gamma & \delta & \delta
\end{array}
\qquad
\begin{array}{c|c}
\mathfrak{A} & \overline{\mathfrak{A}} \\
\hline
\alpha & \delta \\
\beta & \gamma \\
\gamma & \beta \\
\delta & \alpha
\end{array}$$

$(E\,x)\,\mathfrak{A}(x) =$ Erster Wert $w$ in der Wertanordnung $\alpha, \beta, \gamma, \delta$, für den es ein $x$ mit $\mathfrak{A}(x) = w$ gibt.

$(x)\,\mathfrak{A}(x) =$ Letzter Wert $w$ in der Wertanordnung $\alpha, \beta, \gamma, \delta$, für den es ein $x$ mit $\mathfrak{A}(x) = w$ gibt.

Die Umsetzungen sind erfüllt bis auf die *Assoziativität*. Es ist nämlich

$$\alpha \vee (\alpha \vee \gamma) = \alpha \vee \beta = \alpha,$$
$$(\alpha \vee \alpha) \vee \gamma = \alpha \vee \gamma = \beta.$$

4)

$$\begin{array}{cc|cccc}
 & \mathfrak{A} \vee \mathfrak{B} & \alpha & \beta & \gamma & \delta \\
\hline
\mathfrak{A} & \alpha & \alpha & \alpha & \gamma & \gamma \\
 & \beta & \alpha & \beta & \gamma & \gamma \\
 & \gamma & \gamma & \gamma & \gamma & \gamma \\
 & \delta & \gamma & \gamma & \gamma & \delta
\end{array}
\qquad
\begin{array}{cc|cccc}
 & \mathfrak{A} \mathbin{\&} \mathfrak{B} & \alpha & \beta & \gamma & \delta \\
\hline
\mathfrak{A} & \alpha & \alpha & \beta & \beta & \beta \\
 & \beta & \beta & \beta & \beta & \beta \\
 & \gamma & \beta & \beta & \gamma & \delta \\
 & \delta & \beta & \beta & \delta & \delta
\end{array}
\qquad
\begin{array}{c|c}
\mathfrak{A} & \overline{\mathfrak{A}} \\
\hline
\alpha & \delta \\
\beta & \gamma \\
\gamma & \beta \\
\delta & \alpha
\end{array}$$

Quantorenwertung wie unter 3).

Die Umsetzungen sind erfüllt bis auf die *Distributivität*. Diese ist durchbrochen z. B. bei

$$(\alpha \mathbin{\&} \gamma) \vee \alpha = \beta \vee \alpha = \alpha,$$
$$(\alpha \vee \alpha) \mathbin{\&} (\gamma \vee \alpha) = \alpha \mathbin{\&} \gamma = \beta.$$

Bei den letzten beiden Wertungen sind nur Formeln mit den Werten $\alpha$, $\beta$, $\gamma$ herleitbar.

Damit sind die Unabhängigkeitsnachweise für alle Umsetzungsschemata erbracht.

## § 6. Die Normalform einer Herleitung in $K_2$.

Die Herleitungen in $K_2$ lassen sich bezüglich der Reihenfolge ihrer Schlüsse in eine bestimmte Normalform bringen.

Zunächst kann man sich auf diejenigen Disjunktionsschlüsse

$$\frac{\mathfrak{A}}{\mathfrak{A} \vee \mathfrak{B}}$$

beschränken, bei denen das hinzutretende Glied $\mathfrak{B}$ entweder eine Primformel oder die Negation einer Primformel ist. Man verlegt nämlich die Negationszeichen in $\mathfrak{B}$ durch Umsetzungen nach innen bis über die Primbestandteile und streicht gemäß dem Schema der doppelten Negation doppelt auftretende Negationszeichen. Die so umgewandelte Formel $\mathfrak{B}$ setzt sich aus Primbestandteilen und negierten Primbestandteilen allein mittels der Zeichen $\vee$, $\&$, $(E.)$, $(.)$ zusammen. Ersetzt man nun immer die Disjunktionsschlüsse der Formen

$$\frac{\mathfrak{A}}{\mathfrak{A} \vee \mathfrak{B}_1 \vee \mathfrak{B}_2} \qquad \frac{\mathfrak{A}}{\mathfrak{A} \vee (\mathfrak{B}_1 \,\&\, \mathfrak{B}_2)} \qquad \frac{\mathfrak{A}}{\mathfrak{A} \vee (E\,x)\,\mathfrak{B}(x)} \qquad \frac{\mathfrak{A}}{\mathfrak{A} \vee (x)\,\mathfrak{B}(x)}$$

durch die Schlußfiguren

$$\frac{\dfrac{\mathfrak{A}}{\mathfrak{A} \vee \mathfrak{B}_1}}{\mathfrak{A} \vee \mathfrak{B}_1 \vee \mathfrak{B}_2} \qquad \begin{array}{c} \dfrac{\dfrac{\mathfrak{A}}{\mathfrak{A} \vee \mathfrak{B}_1} \quad \dfrac{\mathfrak{A}}{\mathfrak{A} \vee \mathfrak{B}_2}}{(\mathfrak{A} \vee \mathfrak{B}_1) \,\&\, (\mathfrak{A} \vee \mathfrak{B}_2)} \\ \mathfrak{A} \vee (\mathfrak{B}_1 \,\&\, \mathfrak{B}_2) \end{array} \qquad \begin{array}{c} \dfrac{\dfrac{\mathfrak{A}}{\mathfrak{A} \vee \mathfrak{B}(a)}}{(E\,x)\,[\mathfrak{A} \vee \mathfrak{B}(x)]} \\ \mathfrak{A} \vee (E\,x)\,\mathfrak{B}(x) \end{array} \qquad \begin{array}{c} \dfrac{\dfrac{\mathfrak{A}}{\mathfrak{A} \vee \mathfrak{B}(a)}}{(x)\,[\mathfrak{A} \vee \mathfrak{B}(x)]} \\ \mathfrak{A} \vee (x)\,\mathfrak{B}(x) \end{array}$$

so werden schließlich die Disjunktionsschlüsse in der gewünschten Weise reduziert.

Sodann verlegen wir alle Disjunktionsschlüsse über die anderen Schlüsse nach oben. Steht nämlich über einem Disjunktionsschluß

$$\frac{\mathfrak{C}}{\mathfrak{C} \vee \mathfrak{D}}\ (\mathrm{a})$$

ein Schluß

$$\frac{\mathfrak{A} \quad \mathfrak{B}}{\mathfrak{A} \,\&\, \mathfrak{B}}\ (\mathrm{b}), \qquad \frac{\mathfrak{A}(\mathfrak{t})}{(E\,x)\,\mathfrak{A}(x)}\ (\mathrm{c}) \quad \text{oder} \quad \frac{\mathfrak{A}(a)}{(x)\,\mathfrak{A}(x)}\ (\mathrm{d}),$$

wobei die Oberformel $\mathfrak{C}$ des Disjunktionsschlusses durch Umsetzungen aus der Unterformel des darüber stehenden Schlusses hervorgeht, so ersetzen wir den Herleitungsteil durch

$$\begin{array}{c} \dfrac{\dfrac{\mathfrak{A}}{\mathfrak{A} \vee \mathfrak{D}}\,(\mathrm{a}) \quad \dfrac{\mathfrak{B}}{\mathfrak{B} \vee \mathfrak{D}}\,(\mathrm{a})}{(\mathfrak{A} \vee \mathfrak{D}) \,\&\, (\mathfrak{B} \vee \mathfrak{D})}\,(\mathrm{b}) \\ (\mathfrak{A} \,\&\, \mathfrak{B}) \vee \mathfrak{D} \\ \mathfrak{C} \vee \mathfrak{D} \end{array} \qquad \begin{array}{c} \dfrac{\dfrac{\mathfrak{A}(\mathfrak{t})}{\mathfrak{A}(\mathfrak{t}) \vee \mathfrak{D}}\,(\mathrm{a})}{(E\,x)\,[\mathfrak{A}(x) \vee \mathfrak{D}]}\,(\mathrm{c}) \\ (E\,x)\,\mathfrak{A}(x) \vee \mathfrak{D} \\ \mathfrak{C} \vee \mathfrak{D} \end{array} \quad \text{bzw.} \quad \begin{array}{c} \dfrac{\dfrac{\mathfrak{A}(a)}{\mathfrak{A}(a) \vee \mathfrak{D}}\,(\mathrm{a})}{(x)\,[\mathfrak{A}(x) \vee \mathfrak{D}]}\,(\mathrm{d}) \\ (x)\,\mathfrak{A}(x) \vee \mathfrak{D} \\ \mathfrak{C} \vee \mathfrak{D} \end{array}$$

Im letzten Falle ist die freie Variable $a$ in $\mathfrak{A}(a)$ und dem zugehörigen Herleitungsteil eventuell so umzubenennen, daß die Variable nicht in $\mathfrak{D}$ auftritt. Durch die angegebenen Ersetzungen können schrittweise alle Disjunktionsschlüsse über die anderen Schlüsse hinweggezogen werden.

Ferner werden alle Konjunktionsschlüsse über die Quantorenschlüsse (Existenz- und Allschlüsse) verlegt. Man ersetzt nämlich die Schlußfolgen

$$\begin{array}{c}\dfrac{\mathfrak{A}(a)}{(x)\,\mathfrak{A}(x)}\,(\mathrm{d})\\ \dfrac{\mathfrak{C}\qquad\mathfrak{D}}{\mathfrak{C}\,\&\,\mathfrak{D}}\,(\mathrm{b})\end{array}\quad\text{durch}\quad\begin{array}{c}\dfrac{\dfrac{\mathfrak{A}(a)\qquad\mathfrak{D}}{\mathfrak{A}(a)\,\&\,\mathfrak{D}}\,(\mathrm{b})}{(x)[\mathfrak{A}(x)\,\&\,\mathfrak{D}]}\,(\mathrm{d})\\ (x)\,\mathfrak{A}(x)\,\&\,\mathfrak{D}\\ \mathfrak{C}\,\&\,\mathfrak{D}\end{array}$$

Dabei ist die Variable $a$ wieder so zu wählen, daß sie in $\mathfrak{D}$ nicht auftritt. Mit den Existenzschlüssen und den Quantorenschlüssen über rechten Oberformeln von Konjunktionsschlüssen ist entsprechend zu verfahren.

Schließlich lassen sich noch alle Umsetzungen an den Schluß einer Herleitung verlegen, da ja die an den Oberformeln eines Schlusses vorgenommenen Umsetzungen ebenso auf die Unterformeln übertragen werden können. Das ist hier deshalb möglich, weil in den Schlußschematen an die Struktur der Oberformeln gar keine Anforderungen gestellt werden (bis auf die Variablenbedingung, welche sich aber nur auf die Argumente und nicht auf die äußere logische Form bezieht).

Hiernach läßt sich für jede Herleitung in $K_2$ eine *Normalform* folgender Art aufstellen:

1. *Umsetzungen* finden nur unmittelbar über der Endformel $\mathfrak{E}$ statt, welche hierdurch aus einer *pränexen Normalformel* $\mathfrak{E}^*$ (d. h. einer Formel mit voranstehenden Quantoren und nachfolgender konjunktiver Normalform) entsteht.

2. Von $\mathfrak{E}^*$ führt ein unverzweigter Herleitungsfaden allein über *Quantorenschlüsse* nach oben zu einer *aussagenlogischen Formel* $\mathfrak{E}^{**}$ (einer konjunktiven Normalform, welche keine Quantoren enthält).

3. Oberhalb $\mathfrak{E}^{**}$ schließen sich Gabelungen der Herleitung in *Konjunktionsschlüssen* an, durch welche die Formel $\mathfrak{E}^{**}$ aus ihren Primärbestandteilen zusammengesetzt wird.

4. Von den Primärformeln an den oberen Enden der Konjunktionsschlüsse verlaufen unverzweigte Herleitungsfäden allein über *Disjunktionsschlüsse* nach oben bis zu den *Grundformeln*. Die in den Disjunktionsschlüssen hinzutretenden Formelglieder sind ausnahmslos Primformeln oder Negationen von Primformeln.

Die Herleitbarkeit der Formel $\mathfrak{E}^*$ ist auf Grund ihrer Struktur trivial. Läßt man nämlich hier die voranstehenden Quantoren fort und ersetzt die gebundenen Variablen beliebig durch freie Variablen, so hat man eine konjunktive Normalform, deren Primärbestandteile je eine Grundformel enthalten. Untersuchungen über die Herleitbarkeit einer Formel haben also in $K_2$ ihren Schwerpunkt bei den *Umsetzungen*. Es kommt darauf an, eine vorgelegte Formel mittels der Umsetzungsschemata so umzusetzen, daß ihre Herleitbarkeit (oder Nichtherleitbarkeit) offenbar wird. Dabei tritt die algebraische Natur der Logik in den Vordergrund.

Es ist zu beachten, daß die Normaldarstellung in gleicher Weise für Erweiterungen von $K_2$ durch abgeschlossene Axiomensysteme von Primärformeln gilt. Es sind dann nur die Axiome zu den Grundformeln hinzuzunehmen.

## § 7. Der intuitionistische Kalkül $K_3$.

In $K_1$ und $K_2$ wird die vollständige Prädikatenlogik direkt entwickelt, ohne daß Teilbereiche wie die intuitionistische Logik oder der Minimalkalkül abgesondert werden können. Für diese Bereiche ist nämlich eine

andere Wahl von logischen Grundverknüpfungen erforderlich. Wir bilden deshalb für die intuitionistische Logik gesondert einen genetischen Kalkül $K_3$. Als neue Grundzeichen treten hier das Zeichen für die

*Implikation* $\mathfrak{A} \to \mathfrak{B}$ („wenn $\mathfrak{A}$, so $\mathfrak{B}$")

und das Zeichen

$\curlywedge$ für die *falsche Aussage*

auf. Die logischen Grundzeichen von $K_3$ sind

$$\to,\ \vee,\ \&,\ (Ex),\ (x),\ \curlywedge.$$

Die *Negation* wird nicht als Grundverknüpfung behandelt, sondern kann explizit durch

$$\overline{\mathfrak{A}} \equiv \mathfrak{A} \to \curlywedge$$

definiert werden.

Zur Ersparung von Klammern setzen wir fest, daß ein Ausdruck der Form

$$\mathfrak{A}_1 \to \mathfrak{A}_2 \to \cdots \to \mathfrak{A}_n \to \mathfrak{B}$$

als von rechts geklammert anzusehen ist, also die Formel

$$\mathfrak{A}_1 \to (\mathfrak{A}_2 \to (\cdots \to (\mathfrak{A}_n \to \mathfrak{B})\,.\,.))$$

bedeutet. Wir bezeichnen die $\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_n$ als „*Vorderglieder*" der betreffenden Formel und das Glied $\mathfrak{B}$, falls es keine Implikation ist, als das „*Endglied*". Ist dagegen $\mathfrak{B}$ eine Implikation, so ist zur Bestimmung der Vorderglieder und des Endgliedes die Zerlegung von $\mathfrak{B}$ fortzuführen. Jede Formel setzt sich offenbar eindeutig aus ihren Vordergliedern und ihrem Endglied zusammen[8]). Die Vorderglieder können auch fehlen, aber nicht das Endglied. Dieses ist stets entweder eine Primformel (wozu auch das Zeichen $\curlywedge$ gehört) oder hat als äußerstes logisches Zeichen das Disjunktionszeichen, Konjunktionszeichen, Existenz- oder Allzeichen. Reihen von Vordergliedern bezeichnen wir mit großen griechischen Buchstaben. Bedeutet z. B. $\Gamma$ die Reihe $\mathfrak{C}_1, \mathfrak{C}_2, \ldots, \mathfrak{C}_n$, so steht

$$\mathfrak{A} \to \Gamma \to \mathfrak{B}$$

für die Formel

$$\mathfrak{A} \to \bigl(\mathfrak{C}_1 \to (\mathfrak{C}_2 \to (\cdots(\mathfrak{C}_n \to \mathfrak{B})\,.\,.))\bigr).$$

Der Kalkül $K_3$ besitzt die *Grundformeln*

$$\left.\begin{array}{ll} \alpha) & \mathfrak{P} \to \mathfrak{P} \\ \beta) & \curlywedge \to \mathfrak{P} \end{array}\right\} \text{(mit Primformeln } \mathfrak{P})$$

Die Schlußschemata sind

*I. Umformende Schlußschemata*

a) $\dfrac{\Gamma \to \mathfrak{A} \to \mathfrak{B} \to \mathfrak{C}}{\Gamma \to \mathfrak{B} \to \mathfrak{A} \to \mathfrak{C}}$ (Vertauschungen)

b) $\dfrac{\mathfrak{A} \to \mathfrak{A} \to \mathfrak{C}}{\mathfrak{A} \to \mathfrak{C}}$ (Zusammenziehungen)

---

[8]) Hier besteht eine stärkere Analogie zum Gentzenschen Sequenzenkalkül als bei $K_1$ und $K_2$. Die Vorderglieder entsprechen den Antezedensformeln, das Endglied dem Sukzedens (welches ja im intuitionistischen Fall nur eine Formel enthält).

*II. Aufbauende Schlüsse*

1. der Implikation

a) $\dfrac{\mathfrak{C}}{\mathfrak{A} \rightarrow \mathfrak{C}}$ b) $\dfrac{\Gamma \rightarrow \mathfrak{A} \quad \mathfrak{B} \rightarrow \mathfrak{C}}{\Gamma \rightarrow (\mathfrak{A} \rightarrow \mathfrak{B}) \rightarrow \mathfrak{C}}$

2. der Disjunktion

a) $\dfrac{\Gamma \rightarrow \mathfrak{A}}{\Gamma \rightarrow \mathfrak{A} \vee \mathfrak{B}} \quad \dfrac{\Gamma \rightarrow \mathfrak{B}}{\Gamma \rightarrow \mathfrak{A} \vee \mathfrak{B}}$ b) $\dfrac{\mathfrak{A} \rightarrow \mathfrak{C} \quad \mathfrak{B} \rightarrow \mathfrak{C}}{\mathfrak{A} \vee \mathfrak{B} \rightarrow \mathfrak{C}}$

3. der Konjunktion

a) $\dfrac{\mathfrak{A} \rightarrow \mathfrak{C}}{\mathfrak{A} \,\&\, \mathfrak{B} \rightarrow \mathfrak{C}} \quad \dfrac{\mathfrak{B} \rightarrow \mathfrak{C}}{\mathfrak{A} \,\&\, \mathfrak{B} \rightarrow \mathfrak{C}}$ b) $\dfrac{\Gamma \rightarrow \mathfrak{A} \quad \Gamma \rightarrow \mathfrak{B}}{\Gamma \rightarrow \mathfrak{A} \,\&\, \mathfrak{B}}$

4. der Existenzformeln

a) $\dfrac{\Gamma \rightarrow \mathfrak{A}(\mathfrak{t})}{\Gamma \rightarrow (E\,x)\,\mathfrak{A}(x)}$ b) $\dfrac{\mathfrak{A}(a) \rightarrow \mathfrak{C}}{(E\,x)\,\mathfrak{A}(x) \rightarrow \mathfrak{C}}$
(mit Variablenbedingung)

5. der Allformeln

a) $\dfrac{\mathfrak{A}(\mathfrak{t}) \rightarrow \mathfrak{C}}{(x)\,\mathfrak{A}(x) \rightarrow \mathfrak{C}}$ b) $\dfrac{\Gamma \rightarrow \mathfrak{A}(a)}{\Gamma \rightarrow (x)\,\mathfrak{A}(x)}$
(mit Variablenbedingung)

*III. Schnittschema*

$$\frac{\Gamma \rightarrow \mathfrak{A} \quad \mathfrak{A} \rightarrow \mathfrak{C}}{\Gamma \rightarrow \mathfrak{C}}$$

Die mit $\Gamma$ bezeichneten Formelreihen können auch leer sein. In diesem Falle bedeutet z. B. $\Gamma \rightarrow \mathfrak{A}$ die Formel $\mathfrak{A}$. Diese Formelreihen und die mit $\mathfrak{C}$ bezeichneten Formeln bilden die „*Nebenglieder*", die übrigen die „*Hauptglieder*" der betreffenden Schlüsse. Das Hauptglied eines Schnittes ist das „*Schnittglied*". Die Nebenglieder $\mathfrak{C}$ können außer einem Endglied auch Vorderglieder in sich enthalten.

Die Schlußschemata IIa stellen *Abschwächungen, die* Schemata IIb (außer II 1b) *Verstärkungen* dar.

Der intuitionistische Kalkül $K_3$ reduziert sich bei Fortfall der Grundformeln $\beta$) auf den *Minimalkalkül.* Weitere Abgrenzungen sind durch Beschränkung der logischen Zeichen und der entsprechenden aufbauenden Schlußschemata möglich. Man braucht zunächst nur das Implikationszeichen zuzulassen, wobei die aufbauenden Schlüsse auf II 1 beschränkt bleiben, und kann dann in irgendeiner Reihenfolge weitere logische Grundzeichen mit den entsprechenden aufbauenden Schlußschematen hinzunehmen. Wir werden sehen, daß für alle diese Bereiche die Schnitte eliminierbar sind.

Zunächst stellen wir noch fest, daß die Verstärkungen unter den aufbauenden Schlußschematen, nämlich II 2b, II 3b, II 4b, II 5b in $K_3$ *umkehrbar* sind. Der Beweis verläuft ebenso wie der entsprechende für $K_1$. Wir betrachten in der Herleitung einer Formel

(1) $\mathfrak{A} \vee \mathfrak{B} \rightarrow \mathfrak{C}, \quad \Gamma \rightarrow \mathfrak{A} \,\&\, \mathfrak{B}, \quad (E\,x)\,\mathfrak{A}(x) \rightarrow \mathfrak{C}$ bzw. $\Gamma \rightarrow (x)\,\mathfrak{A}(x)$

den Formelbund von

(2) $\mathfrak{A} \vee \mathfrak{B}, \quad \mathfrak{A} \,\&\, \mathfrak{B}, \quad (E\,x)\,\mathfrak{A}(x)$ bzw. $(x)\,\mathfrak{A}(x)$.

Dieser führt durch Nebenglieder von aufbauenden Schlüssen und Schnitten nach oben entweder zu Hauptgliedern von II 1a-Schlüssen (Implikations-Abschwächungen) oder zu Hauptgliedern der betrachteten Verstärkungs-Schlüsse. Es ist nämlich zu beachten, daß der Formelbund eines Gliedes (2) immer nur aus Vordergliedern oder immer nur aus Endgliedern von Formeln der Herleitung besteht. Es können nun wie bei $K_1$ aus Herleitungen für Formeln (1) allein durch Ersetzungen von Formelbestandteilen und eventuelle Streichungen von Herleitungsteilen Herleitungen für

$$\mathfrak{A} \to \mathfrak{C},\ \mathfrak{B} \to \mathfrak{C};\quad \Gamma \to \mathfrak{A},\ \Gamma \to \mathfrak{B};\quad \mathfrak{A}(a) \to \mathfrak{C} \text{ bzw. } \Gamma \to \mathfrak{A}(a)$$

gewonnen werden.

## § 8. Elimination der Schnitte in $K_3$.

Um die Eliminierbarkeit der Schnitte in $K_3$ zu beweisen, denken wir uns wieder eine Herleitung gegeben, die einen obersten Schnitt

$$(\Sigma)\qquad \frac{P \to \mathfrak{S} \cdot\ \ \mathfrak{S} \to \mathfrak{T}}{P \to \mathfrak{T}}$$

enthält. Das Verfahren zur Beseitigung des Schnittes entspricht dem von $K_1$, ist nur etwas auszugestalten.

*I. Fall.* Das Schnittglied $\mathfrak{S}$ sei eine Primformel.

I a) Oberhalb der rechten Oberformel mögen keine aufbauenden Schlüsse auftreten.

Dann ist $\mathfrak{S} \to \mathfrak{T}$ eine Grundformel (da sich an Grundformeln keine umformenden Schlüsse anschließen können). Ist $\mathfrak{S} \to \mathfrak{T}$ eine Grundformel $\alpha$), so stimmt $\mathfrak{S}$ mit $\mathfrak{T}$ überein, und der Schnitt kann fortfallen. Ist $\mathfrak{S} \to \mathfrak{T}$ eine Grundformel $\beta$), so lautet die linke Oberformel $P \to \curlywedge$, und es ist $\mathfrak{T}$ eine Primformel. Der Formelbund von $\curlywedge$ über der linken Oberformel gehört nur zu Nebengliedern (Endgliedern) von umformenden und aufbauenden Schlüssen sowie zu Endgliedern von Grundformeln $\curlywedge \to \curlywedge$. Wir ersetzen alle Glieder des Bundes durch die Primformel $\mathfrak{T}$, wodurch die Grundformeln des Bundes in Grundformeln $\beta$) übergehen. Dabei bleibt der Herleitungszusammenhang gewahrt, und aus der linken Oberformel wird die Formel $P \to \mathfrak{T}$, welche mit der Unterformel des Schnittes übereinstimmt. Der Schnitt kann also fortfallen.

I b) Oberhalb der rechten Oberformel möge mindestens ein aufbauender Schluß auftreten. Die dem Schnittglied zugehörigen Glieder gehören dem letzten aufbauenden Schluß über $\mathfrak{S} \to \mathfrak{T}$ entweder als Nebenglieder oder als Hauptglieder von Implikations-Abschwächungen an. In jedem Falle läßt sich wie bei $K_1$ der Schnitt über den letzten aufbauenden Schluß hinwegziehen (bzw. ganz beseitigen).

*II. Fall.* Das Schnittglied $\mathfrak{S}$ sei von der Form $\mathfrak{A} \to \mathfrak{B}$. Der hierzu gehörende Formelbund über der rechten Oberformel $(\mathfrak{A} \to \mathfrak{B}) \to \mathfrak{T}$ enthält $\mathfrak{A} \to \mathfrak{B}$ entweder als Nebenglied (Vorderglied) von aufbauenden Schlüssen oder als Hauptglied von Implikations-Abschwächungen oder drittens als Hauptglied in Unterformeln

$$\Gamma_i \to (\mathfrak{A} \to \mathfrak{B}) \to \mathfrak{C}_i$$

von II 1b-Schlüssen mit den Oberformeln

$$\Gamma_i \to \mathfrak{A}, \quad \mathfrak{B} \to \mathfrak{C}_i .$$

Unter Verwendung der linken Oberformel von $(\Sigma)$ $P \to \mathfrak{A} \to \mathfrak{B}$ ersetzen wir diese II 1b-Schlüsse überall durch

$$\begin{array}{ll} \dfrac{P \to \mathfrak{A} \to \mathfrak{B} \quad \mathfrak{B} \to \mathfrak{C}_i}{P \to \mathfrak{A} \to \mathfrak{C}_i} & \text{(Schnitt mit Schnittglied } \mathfrak{B}\text{)} \\ & \text{(Vertauschungen)} \\ \dfrac{\Gamma_i \to \mathfrak{A} \quad \mathfrak{A} \to P \to \mathfrak{C}_i}{\Gamma_i \to P \to \mathfrak{C}_i} & \text{(Schnitt mit Schnittglied } \mathfrak{A}\text{)} \end{array}$$

Ersetzen wir weiterhin alle Glieder $\mathfrak{A} \to \mathfrak{B}$ des Formelbundes durch die Reihe $P$, so geht unter Wahrung des Herleitungszusammenhangs die rechte Oberformel von $(\Sigma)$ in $P \to \mathfrak{T}$ über, so daß der Schnitt fortfällt.

*III. Fall.* Das Schnittglied $\mathfrak{S}$ sei von einer der Formen

$$\mathfrak{A} \vee \mathfrak{B}, \quad \mathfrak{A} \,\&\, \mathfrak{B}, \quad (E\,x)\,\mathfrak{A}(x), \quad (x)\,\mathfrak{A}(x).$$

Auf Grund der Umkehrbarkeit der aufbauenden Verstärkungsschlüsse gewinnen wir aus der Herleitung der rechten Oberformel (im 1. und 3. Fall) bzw. der linken Oberformel (im 2. und 4. Fall) schnittfreie Herleitungen von

$$\mathfrak{A} \to \mathfrak{T},\ \mathfrak{B} \to \mathfrak{T}; \quad P \to \mathfrak{A},\ P \to \mathfrak{B}; \quad \mathfrak{A}(a) \to \mathfrak{T} \text{ bzw. } P \to \mathfrak{A}(a).$$

Bringen wir diese Formeln entsprechend wie bei $K_1$ zum Schnitt mit Herleitungsteilen der anderen Oberformel, so gewinnen wir eine Herleitung von $P \to \mathfrak{T}$ unter Beseitigung des Schnittes $(\Sigma)$. An dessen Stelle sind ebenso wie im Falle II Schnitte getreten, deren Schnittglieder weniger logische Zeichen als das Schnittglied von $(\Sigma)$ enthalten.

Es lassen sich also wie bei $K_1$ durch wiederholte Anwendung des angegebenen Verfahrens schließlich alle Schnitte beseitigen. Aus dem Ersetzungsverfahren ist zu ersehen, daß die Schnitte auch im Minimalkalkül (bei Fortfall der Grundformeln $\beta$)) und in den Teilbereichen bei Beschränkung auf weniger logische Verknüpfungen entbehrlich sind.

Ein Entscheidungsverfahren für die intuitionistische Aussagenlogik ist auf Grund der Eliminierbarkeit der Schnitte für $K_3$ ebenso zu gewinnen, wie dies GENTZEN[9]) für den Sequenzenkalkül gezeigt hat.

---

[9]) Vgl. die in Anmerkung [1]) zitierte Arbeit.

*(Eingegangen am 31. Juli 1949.)*

Math. Annalen, Bd. 122, S. 66—70 (1950).

# Über analytische Automorphismen des $R_{2n}$.

Von

HANS HERMES in Münster (Westf.) und ERNST PESCHL in Bonn.

1. Unter einem *Automorphismus* des $R_{2n}$ soll im folgenden eine umkehrbar eindeutige und analytische Abbildung des Raumes $R_{2n}$ von $n$ komplexen Dimensionen auf sich verstanden werden. Jede solche Abbildung ist darstellbar in der Form:

$$x'_j = f_j(x_1, \ldots, x_n) \qquad (j = 1, \ldots, n),$$

wobei die $f_j$ ganze Funktionen der komplexen Variablen $x_1, \ldots, x_n$ sind.

Man übersieht die Menge dieser Automorphismen bis jetzt nicht vollständig. Bekannt sind jedoch folgende Grundtypen, aus denen sich leicht allgemeinere Automorphismen zusammensetzen lassen:

A. Die *linearen Transformationen:*

$$x'_j = \sum_{k=1}^{n} a_{jk}\, x_k \qquad (j = 1, \ldots, n)$$

mit nicht verschwindender Determinante $|a_{jk}|$.

B. Die *Schiebungen*, die man bei Auszeichnung der ersten Koordinate so schreiben kann:

$$\begin{aligned} x'_1 &= x_1 \\ x'_j &= x_j + f_j(x_1), \quad f_j \text{ ganze Funktionen } (j = 2, \ldots, n), \end{aligned}$$

mit der Umkehrung:

$$\begin{aligned} x_1 &= x'_1 \\ x_j &= x'_j - f_j(x'_1) \quad (j = 2, \ldots, n). \end{aligned}$$

C. Automorphismen der Form:

$$x'_j = x_j\, e^{r_j f\left(x_1^{s_1} \ldots x_n^{s_n}\right)} \qquad (j = 1, \ldots, n).$$

Hierbei soll $f(z)$ eine ganze Funktion sein; die $r_j$ seien ganze und die $s_j$ nicht negative ganze Zahlen, die der Nebenbedingung:

$$\sum_{j=1}^{n} r_j\, s_j = 0$$

genügen. Man verifiziert leicht:

$$x_1'^{s_1} \ldots x_n'^{s_n} = x_1^{s_1} \ldots x_n^{s_n}.$$

Also hat ein solcher Automorphismus die Umkehrung:

$$x_j = x'_j\, e^{-r_j f\left(x_1'^{s_1} \ldots x_n'^{s_n}\right)} \qquad (j = 1, \ldots n).$$

2. Die Punktfolgen $P_\nu = \{(x_{1\nu}, \ldots, x_{n\nu})\}$ und $P'_\nu = \{(x'_{1\nu}, \ldots, x'_{n\nu})\}$ des $R_{2n}$ sollen *äquivalent* heißen, wenn es einen Automorphismus des $R_{2n}$ gibt,

der für jedes $\nu$ $P_\nu$ in $P'_\nu$ überführt. Die Folge $P_\nu$ heiße *plan*, wenn alle $P_\nu$ in einer zweidimensionalen analytischen Ebene liegen. Eine solche Ebene läßt sich geben in der Parameterdarstellung:

$$x_j = a_j\, t + b_j \qquad (j = 1, \ldots, n),$$

wobei wenigstens ein $a_j$ nicht verschwindet. Eine zu einer planen Folge äquivalente Folge soll *planierbar* heißen.

In Nr. 3 soll für gewisse Punktfolgen die Planierbarkeit bewiesen werden. Von vorneherein sollen jedoch nur solche Punktfolgen betrachtet werden, die nur verschiedene Glieder haben, und im Endlichen keinen Häufungspunkt. *Diese Voraussetzungen werden von nun an immer gemacht, auch wenn sie nicht explizit erwähnt werden.* Eine solche Folge ist endlich oder abzählbar. Im Spezialfall der Endlichkeit sind die behandelten Fragen trivial, so daß von jetzt ab eine solche Folge stets als abzählbar vorausgesetzt werden soll.

Die Beweise verwenden die in Nr. 1 aufgezählten speziellen Automorphismen in Verbindung mit der folgenden bekannten Tatsache (Lösbarkeit des Interpolationsproblems für eine Variable): Es sei im $R_2$ eine Folge $\{x_\nu\}$ verschiedener Punkte ohne Häufungspunkt im Endlichen vorgegeben. Dann gibt es zu beliebigen Werten $a_\nu$ stets eine ganze Funktion $f$, die die Bedingungen $f(x_\nu) = a_\nu$ für jedes $\nu$ erfüllt[1]). In Nr. 4 wird gezeigt, daß das analoge Interpolationsproblem auch im $R_{2n}$ eine Lösung besitzt.

3. Die Punktfolgen des $R_{2n}$ zerfallen in Klassen äquivalenter Folgen. Es bleibt hier unentschieden, ob es mehr als eine Klasse gibt. Diese Frage kann aber reduziert werden auf die Diskussion des Problems, ob alle Folgen planierbar sind. Es gilt nämlich der

*Satz 1*: Alle planierbaren Folgen sind äquivalent.

*Beweis:* Es genügt, zu zeigen, daß zwei plane Folgen $P_\nu = \{(x_{1\nu}, \ldots, x_{n\nu})\}$ und $P'_\nu = \{(x'_{1\nu}, \ldots, x'_{n\nu})\}$ äquivalent sind. Die $P_\nu$ mögen auf der analytischen Ebene $E$ mit der Parameterdarstellung $x_j = a_j\, t + b_j$ liegen, analog die $P'_\nu$ auf der Ebene $E'$ mit der Parameterdarstellung $x_j = a'_j\, t + b'_j$. Es gibt ein $k$, so daß $a_k \neq 0$. Also sind die $\{x_{k\nu}\}$ alle verschieden und haben keinen endlichen Häufungspunkt. Ebenso gibt es ein $k'$, so daß die $\{x'_{k'\nu}\}$ alle verschieden sind und keinen endlichen Häufungspunkt haben. Es kann angenommen werden, daß $k \neq k'$; sonst läßt sich dies durch einen achsenvertauschenden linearen Automorphismus erreichen. Nach dem in Nr. 1 genannten Interpolationssatz gibt es ganze Funktionen $f$ und $g_j$, die für jedes $\nu$ den Bedingungen:

$$f(x_{k\nu}) = x'_{k'\nu} - x_{k'\nu}$$

$$\text{bzw.}\quad g_j(x'_{k'\nu}) = x'_{j\nu} - x_{j\nu} \qquad (j = 1, \ldots, n)$$

genügen. Man bilde die Automorphismen ($\delta_{jk'}$ ist das Kroneckersymbol):

$$T_1\colon \qquad x'_j = x_j + \delta_{jk'}\, f(x_k) \qquad j = 1, \ldots, n),$$

$$T_2\colon \qquad x''_j = x'_j + (1 - \delta_{jk'})\, g_j(x'_{k'}) \qquad (j = 1, \ldots, n).$$

Man sieht sofort, daß $T_2(T_1)$ die Folge $P_\nu$ in die Folge $P'_\nu$ überführt.

---

[1]) Der Satz stammt von Mittag-Leffler, Acta math. 4, 32 (1884). Siehe z. B. Osgood, Lehrbuch der Funktionentheorie. S. 569—574. Leipzig: C. G. Teubner 1928; oder Knopp, Aufgabensammlung zur Funktionentheorie II, Kap. III, § 10, Aufg. 6 p. 25 (Lösung p. 89) und § 11, Aufg. 4 p. 26 (Lösung p. 92—93). Berlin: W. de Gruyter 1944.

*Korollar*: Eine planierbare Folge $P_\nu$ ist äquivalent zu jeder Folge $P_{\mu_\nu}$, die aus ihr durch Umordnen der Glieder entsteht. Man darf also auch von planierbaren Mengen sprechen.

Nunmehr soll für einige Klassen von Folgen die Planierbarkeit bewiesen werden.

*Satz 2*: Eine Folge $P_\nu = \{(x_{1\nu}, \ldots, x_{n\nu})\}$ ist planierbar, wenn für ein gewisses $j$ die Folge $\{x_{j\nu}\}$ keinen endlichen Häufungspunkt hat.

*Beweis:* O. B. d. A. kann angenommen werden, daß die Folge $\{x_{1\nu}\}$ keinen endlichen Häufungspunkt hat. In a) wird durch Induktion nach $m$ gezeigt: $P_\nu$ ist äquivalent mit einer Folge $P_\nu^{(m)} = \{(x_{1\nu}^{(m)}, \ldots, x_{n\nu}^{(m)})\}$ mit der Eigenschaft, daß die Folge $\{x_{m\nu}^{(m)}\}$ keinen endlichen Häufungspunkt hat, und daß zudem $x_{m\nu}^{(m)}$ nur dann mit $x_{m\mu}^{(m)}$ übereinstimmt, wenn auch $x_{1\nu}^{(m)} = x_{1\mu}^{(m)}, \ldots, x_{m-1,\nu}^{(m)} = x_{m-1,\mu}^{(m)}$. Dann gilt für die Folge $P_\nu^{(n)}$: Die $\{x_{n\nu}^{(n)}\}$ haben keinen endlichen Häufungspunkt und es ist $x_{n\nu}^{(n)} \neq x_{n\mu}^{(n)}$ für $\nu \neq \mu$. Diese Folge $P_\nu^{(n)}$ wird in b) planiert.

a) $P_\nu^{(1)} = P_\nu$ leistet für $m = 1$ das Verlangte. Nachweis der Existenz von $P_\nu^{(m+1)}$: $z_\varrho$ sei eine Aufzählung der Menge aller $x_{m\nu}^{(m)}$. Es sind also alle $z_\varrho$ untereinander verschieden, und es gibt eine Funktion $N(\nu)$, so daß für jedes $\nu$:

$$(*) \qquad x_{m\nu}^{(m)} = z_{N(\nu)}.$$

Weil die $x_{m\nu}^{(m)}$ keinen endlichen Häufungspunkt haben, haben die $z_\varrho$ auch keinen; außerdem folgt, daß die Gleichung $N(\nu) = \varrho$ bei festem $\varrho$ nur endlich viele Lösungen $\nu$ besitzt. Daher lassen sich durch Rekursion nach $\varrho$ Werte $a_\varrho$ finden, für die gilt: Für jedes $\nu$ und $\mu$ folgt aus $N(\nu) < N(\mu)$, daß

$$(**) \qquad N(\nu) < |x_{m+1,\nu}^{(m)} + a_{N(\nu)}| < |x_{m+1,\mu}^{(m)} + a_{N(\mu)}|.$$

Das Interpolationsproblem $f(z_\varrho) = a_\varrho$ besitzt eine ganze Lösung $f$. Der Automorphismus:

$$T: \qquad x_j' = x_j + \delta_{j,m+1} f(x_m) \quad (j = 1, \ldots, n)$$

führe $\{(x_{1\nu}^{(m)}, \ldots, x_{n\nu}^{(m)})\}$ über in $\{(x_{1\nu}^{(m+1)}, \ldots, x_{n\nu}^{(m+1)})\}$. Es ist für jedes $\nu$:

$$x_{m+1,\nu}^{(m+1)} = x_{m+1,\nu}^{(m)} + a_{N(\nu)}.$$

Die $x_{m+1,\nu}^{(m+1)}$ haben wegen (**) keinen endlichen Häufungspunkt. Sei ferner $x_{m+1,\nu}^{(m+1)} = x_{m+1,\mu}^{(m+1)}$. Dann kann wegen (**) nicht $N(\nu) < N(\mu)$ sein. Es bleibt somit nur noch der Fall $N(\nu) = N(\mu)$ zu diskutieren. Dann folgt aber aus (*): $x_{m\nu}^{(m)} = x_{m\mu}^{(m)}$, also wegen der Induktionsvoraussetzung auch $x_{1\nu}^{(m)} = x_{1\mu}^{(m)}, \ldots, x_{m-1,\nu}^{(m)} = x_{m-1,\mu}^{(m)}$; dieselben Gleichungen gelten aber auch, wenn man den oberen Index $(m)$ durch $(m+1)$ ersetzt, da $T$ diese Koordinaten ungeändert läßt.

b) Es gibt ganze Lösungen $g_j$ der Interpolationsprobleme

$$g_j(x_{n\nu}^{(n)}) = -x_{j\nu}^{(n)} \quad (j = 1, \ldots, n).$$

Der Automorphismus:

$$x_j' = x_j + (1 - \delta_{jn})\, g_j(x_n) \quad (j = 1, \ldots, n)$$

planiert offenbar die Folge $P_\nu^{(n)}$.

*Satz 3*: Eine Folge $\{(x_{1\nu}, \ldots, x_{n\nu})\}$ ist planierbar, wenn keine der Folgen $\{x_{j\nu}\}$ den Häufungspunkt 0 hat.

*Beweis*: $z_\varrho$ sei eine Aufzählung der Menge aller Produkte $x_{1\nu} \ldots x_{n\nu}$. Die $z_\varrho$ sind also untereinander verschieden und die Folge $\{z_\varrho\}$ hat keinen endlichen Häufungspunkt. Es sei

$$x_{1\nu} \ldots x_{n\nu} = z_{N(\nu)}.$$

Wegen der Voraussetzung über die $x_{j\nu}$ hat die Gleichung $N(\nu) = \varrho$ für festes $\varrho$ nur endlich viele Lösungen $\nu$. Es gibt eine ganze Funktion $f$ als Lösung des Interpolationsproblems $f(z_\varrho) = \log \varrho \geq 0$. Der Automorphismus:

$$\begin{cases} x'_j = x_j\, e^{r_j f(x_1 \cdots x_n)} & (j = 1, \ldots, n) \\ r_1 = 1, r_2 = -1, r_3 = \cdots = r_n = 0 \end{cases}$$

führe $\{(x_{1\nu}, \ldots, x_{n\nu})\}$ über in $\{(x'_{1\nu}, \ldots, x'_{n\nu})\}$. Wegen $|x'_{1\nu}| = |x_{1\nu}| \cdot N(\nu)$ hat die Folge $\{x'_{1\nu}\}$ keinen endlichen Häufungspunkt. Damit folgt Satz 3 aus Satz 2.

*Korollar*: Jede Folge, die die *Gitterpunkte* des $R_{2n}$ durchläuft (d. h. die Menge der Punkte $(\varrho_1 + i\,\sigma_1, \ldots, \varrho_n + i\,\sigma_n)$ mit ganzen $\varrho, \sigma$), ist planierbar. Vor Anwendung von Satz 3 braucht man eine solche Folge nur abzubilden mittels des Automorphismus:

$$x'_j = x_j + \frac{1}{2} \quad (j = 1, \ldots, n).$$

Die in den Sätzen 2 und 3 gegebenen Kriterien lassen sich noch in verschiedener Richtung verallgemeinern, was aber hier nicht ausgeführt werden soll[2]). Von Interesse ist jedoch der

*Satz 4*: Eine Folge $P_\nu = \{(x_{1\nu}, \ldots, x_{n\nu})\}$ läßt sich stets in $n$ planierbare Teilfolgen aufspalten.

*Beweis*: $P_\nu$ gehöre genau dann zur $m$-ten Teilfolge $P_\nu^{(m)}$ $(m = 1, \ldots, n)$, wenn $|x_{m\nu}|$ von keinem der anderen $|x_{j\nu}|$ übertroffen wird (falls dieses Kriterium nicht eindeutig entscheidet, entscheide man willkürlich). Die $m$-ten Koordinaten der Folge $P_\nu^{(m)}$ haben offenbar keinen endlichen Häufungspunkt. Daher ist Satz 2 anwendbar.

4. Das eingangs formulierte *Interpolationsproblem* für den $R_{2n}$ wird völlig gelöst durch den

*Satz 5*: Sei $\{(x_{1\nu}, \ldots, x_{n\nu})\}$ eine Folge verschiedener Punkte ohne endlichen Häufungspunkt. Dann gibt es zu beliebigen Werten $a_\nu$ eine ganze Funktion $F$ mit $F(x_{1\nu}, \ldots, x_{n\nu}) = a_\nu$[3]).

*Beweis*: Für den $R_2$ ist jedes Interpolationsproblem (d. h. das Interpolationsproblem bei beliebig vorgegebenen Werten $a_\nu$) lösbar (vgl. Nr. 1). Es sei vorausgesetzt, daß jedes Interpolationsproblem für den $R_{2(n-1)}$ lösbar ist. Eine Folge im $R_{2n}$ ist nach Satz 4 in $n$ planierbare Teilfolgen aufspaltbar.

---

[2]) Ohne Beweis sei nur erwähnt, daß in Erweiterung von Satz 2 die Planierbarkeit auch noch bewiesen werden kann, wenn die Folge $\{x_{j\nu}\}$ nur endlich viele endliche Häufungspunkte besitzt.

[3]) Dieser Satz läßt sich im Spezialfall $n = 2$ leicht zurückführen auf einen Interpolationssatz von H. Cartan, Sur les variétés définies par une relation entière, Bull. Sci. math. **55**, 24—32, 47—64 (1931), Théorème I.

Daher kommt man sofort induktiv zum Ziel, wenn man beweist: Läßt sich eine Folge $P_\nu$ des $R_{2n}$ so in zwei Teilfolgen aufspalten, daß für die erste Teilfolge jedes Interpolationsproblem lösbar ist, und daß die zweite Teilfolge planierbar ist, so ist für $P_\nu$ jede Interpolationsaufgabe lösbar. Da ferner sowohl die Lösbarkeit des Interpolationsproblems als auch die Planierbarkeit bei Übergang zu äquivalenten Folgen erhalten bleibt, kann man zudem voraussetzen, daß alle Punkte der zweiten Teilfolge in der Hyperebene $x_n = 0$ liegen. Diejenigen Punkte der ersten Teilfolge, die ebenfalls in dieser Hyperebene liegen, nehme man zur zweiten Teilfolge hinüber. Damit hat man folgendes erreicht: (1) Für beide Teilfolgen ist jedes Interpolationsproblem lösbar (man beachte hierzu, daß die zweite Teilfolge im $R_{2(n-1)}$ liegt); (2) $P_\nu$ gehört zur zweiten bzw. ersten Teilfolge, je nachdem die letzte Koordinate von $P_\nu$ verschwindet oder nicht.

Die Interpolationsaufgabe $F(P_\nu) = F(x_{1\nu}, \ldots, x_{n\nu}) = a_\nu$ kann man nun folgendermaßen lösen: Es gibt eine ganze Funktion $F_1$, so daß

$$F_1(P_\nu) = a_\nu$$

für alle Punkte der zweiten Teilfolge. Weiter gibt es hierzu eine ganze Funktion $F_2$, so daß

$$F_2(P_\nu) = \frac{a_\nu - F_1(P_\nu)}{x_{n\nu}}$$

für alle Punkte der ersten Teilfolge (für die ja $x_{n\nu}$ nicht verschwindet). Dann ist offenbar die ganze Funktion:

$$F = F_1 + x_n F_2$$

eine Lösung des Interpolationsproblems für die gesamte Folge $P_\nu$.

*(Eingegangen am 15. Juni 1949.)*

Math. Annalen, Bd. 122, S. 71—89 (1950).

# Systematische Basisreduktion der Modalitäten bei Idempotenz der positiven Grundmodalitäten.

Von

ARNOLD SCHMIDT in Göttingen.

§ 1. Einleitung.
§ 2. Gemeinsame Grundlage.
§ 3. Modalitätenstrukturen (Definitionen, allgemeine Eigenschaften).
§ 4. Die einstufigen idempotenten Modalitätenstrukturen.
§ 5. Die zweistufigen idempotenten Modalitätenstrukturen.

## § 1. Einleitung.

Die Modalitäten beginnen in der mathematischen Logik eine wachsende Rolle zu spielen. Ursprünglich wohl von MAC COLL (1) in diese eingeführt, sind sie insbesondere einerseits in der Theorie der strikten Implikation von C. I. LEWIS (2,5) und andererseits in der dreiwertigen Logik von J. ŁUKASIEWICZ (3) zu Bedeutung gelangt. Die LEWISsche strikte Implikation wird mit Hilfe der „Möglichkeit" definiert, die als zweite einstellige aussagenlogische Grundoperation zur Negation hinzutritt. O. BECKER (4,9) warf im Rahmen der LEWISschen Aussagenlogik die Fragen nach der Rangordnung und nach der Reduktion der durch Aufeinanderausüben von Möglichkeit und Negation entstehenden Modi auf. Diese Fragen wurden von einer Reihe anderer Autoren weiterbearbeitet; im Zusammenhang mit der folgenden Systematik sind hier wohl insbesondere C. W. CHURCHMAN (7) und W. T. PARRY (8) zu erwähnen, die teils verschiedene, von BECKER vorgeschlagene Modalitätenlogiken als identisch erkannten, teils weitere axiomatische Ergänzungen angaben, mit denen sich die Modi reduzieren lassen (die Arbeit von PARRY (8) ist mir übrigens erst nach der Abfassung des Haupteils (§§ 2—5)[1]) jetzt bekannt geworden; dort herangezogene Arbeiten von R. FEYS (6) waren mir bislang nicht zugänglich).

Nach der Entdeckung einer Reihe von Modalitätenstrukturen scheint nunmehr die Frage nach einer dem Reduktionsprozeß selbst entfließenden Ordnung und einer *vollständigen* Übersicht über die in erster Linie in Frage kommenden Modalitätenstrukturen dringlich zu werden.

Im folgenden soll die Reduzierbarkeit der Modalitäten in *systematischer* Weise untersucht werden. Diese Systematik umfaßt drei Gesichtspunkte. 1. Die Frage der Reduzierbarkeit der Modi stellt sich unabhängig von der Theorie der strikten Implikation, in deren Rahmen sie bisher behandelt zu werden pflegte, und ist von ihren Besonderheiten nicht beeinflußt. Das Problem läßt sich von der begrifflichen Grundlage der bloßen Gleichheiten und Ungleichheiten zwischen Modi aus behandeln. Diese angemessene Grundlage gestattet die allgemein verbindliche Auswertung, die sich dann zu den verschiedenen Theorien der mathematischen Logik (oder auch der Ontologie)

[1]) Die vorliegende Arbeit stammt im wesentlichen aus dem Jahre 1945.

in Beziehung setzen läßt. In der vorliegenden Arbeit beschränke ich mich zunächst auf Gleichheiten. In einer Fortsetzungsarbeit werden sodann die Ungleichheiten einbezogen. (Mit ihnen läßt sich dann auch die Frage nach der Rangordnung der Modalitäten erledigen und somit die Axiomatik der Modalitätenstrukturen vervollständigen.) 2. Ausgehend von den unechten Grundmodi (Zutreffen, Nichtzutreffen) und den echten Grundmodi (Möglichkeit, Unmöglichkeit, Notwendigkeit, Nichtnotwendigkeit) wird man nach Modalitätenstrukturen $n$-ter „Stufe" fragen, bei denen sich alle Produkte von Grundmodalitäten auf eine „Basis" reduzieren, die aus höchstens $n$-elementigen Produkten besteht und ein $n$-elementiges Produkt enthält. Unter der Voraussetzung, daß die echten positiven Grundmodi idempotent seien (möglicherweise möglich = möglich; notwendig = notwendig notwendig), wird die Frage nach den ein- und zweistufigen Modalitätenstrukturen beantwortet, d. h. über die bisher übliche bloße Auffindung von Strukturen hinaus wird die Menge *aller* dieser Strukturen aufgesucht. — 3. Als ein letzter systematischer Gesichtspunkt darf noch der folgende erwähnt werden. Um die in der axiomatischen Grundlegung liegende Willkür auszuschalten und eine etwaige Äquivalenz axiomatisch eingeführter Strukturen nicht der „nachträglichen" Entdeckung zu überlassen, werden für die einzelnen Strukturen bei jedem Schritt der Untersuchung die als mögliche Axiome gleichberechtigten Grundgleichungen — die „Axiomablen" — vollständig aufgeführt. Dadurch tritt zugleich die methodische Symmetrie zwischen den echten Grundwerten — Dualität (von Möglichkeit und Notwendigkeit) und Verneinung, zusammengefaßt zu einer „Quaternalität" — deutlich heraus. (Im Gegensatz zu der in den Absätzen 1., 2. zusammengefaßten Systematik könnte dieser 3. Gesichtspunkt als ein mehr deduktionstechnischer aufgefaßt werden. Ich möchte demgegenüber der Überzeugung Ausdruck geben, daß allgemein eine axiomatische Untersuchung erst dann den mathematischen Gehalt ihres Gebietes ausschöpft, wenn sie jeweils von den sich anbietenden „Axiomablengruppen" ausgeht.)

Die folgenden §§ 2—3 bringen die zu Grunde liegenden präzisen Definitionen und die aus ihnen folgenden grundlegenden Sätze. Die §§ 4—5 sind der Systematik der idempotenten Modalitätenstrukturen 1. und 2. Stufe gewidmet; es ergeben sich *zwei Sechsmodalitätenstrukturen* der 1. Stufe und eine *Achtmodalitätenstruktur*[2]) sowie *drei Zehnmodalitätenstrukturen* der 2. Stufe[3]) (zu den sechs oben genannten Grundmodi treten dabei als unreduzierbare Modi die Modi „widerspruchsfrei" = unmöglich unmöglich = notwendig möglich — und „stringenzfähig" = möglicherweise notwendig = nicht notwendigerweise nichtnotwendig — mit ihren Negationen).

[2]) Die Angabe der konkurrenzfähigen Axiomablen bei jeder neuen Axiomablengruppe bewirkt, daß sich jeweils eine Vielfalt von gleichberechtigten Axiomensystemen anbietet. So werden (auch wenn man von Gleichheiten zwischen negativen Werten absieht, da die Gleichheit zwischen den entsprechenden positiven Werten den Vorzug hat) z. B. für die Achtmodalitätenstruktur der 2. Stufe in § 5 über viertausend (nämlich $4 \cdot 6 \cdot (12 \cdot 12 + 12 \cdot 2) = 4032$ gleichberechtigte Axiomensysteme angeboten.

[3]) In der oben zitierten Arbeit (8) hat PARRY (mit Heranziehung der BECKERschen Ergebnisse (4) die sich im Rahmen der LEWISschen Logik ergebenden Analoga zur 1. Sechs-, zur Acht- und zur 1. Zehnmodalitätenstruktur angegeben, während er das Analogon zur 2. Sechsmodalitätenstruktur als mit der LEWISschen Theorie der strikten Implikation nicht verträglich verwirft. Ein exakter axiomatischer Vergleich ist erst bei Einbeziehung der Ungleichheiten möglich und wird in der angekündigten Fortsetzungsarbeit gegeben werden.

Die erste Zehnmodalitätenstruktur wird durch das Axiom „unmöglich unmöglich unmöglich = unmöglich" beherrscht, das etwa dem Brouwerschen Absurditätssatz entspricht[4]). Auf die Erörterung, welche der formal erhaltenen Strukturen sinnvoll sind, wird in dieser Arbeit verzichtet; die interpretierende Diskussion soll vielmehr an anderer Stelle durchgeführt werden. Es mag jedoch bereits hier hervorgehoben werden, daß keine der sechs erhaltenen idempotenten Modalitätenstrukturen 1. oder 2. Stufe einer Interpretation unfähig ist, wenn auch bei einigen von ihnen die interpretierenden Modalitäten von der üblichen „Möglichkeit" und „Notwendigkeit" abweichen[5]).

## § 2. Gemeinsame Grundlage.

Gegeben seien die Werte $z, m, g$ und $\bar{z}, \bar{m}, \bar{g}$. Diese sechs Zeichen seien einer assoziativen Verknüpfung fähig (m. a. W. die aus diesen Zeichen bestehenden Zeichenreihen sollen eine Halbgruppe bilden).

*Abkürzung:* $n = \bar{z}$.

*Interpretation.* Mit den Grundwerten sind Modalitäten gemeint; die Überstreichung soll Negieren bedeuten:

| | |
|---|---|
| $z$: zutreffend (wahr, wirklich) | $\bar{z}$: unzutreffend (falsch) |
| $m$: möglich (zugelassen) | $\bar{m}$: unmöglich (verboten, absurd, widerspruchsvoll) |
| $g$: gewiß, notwendig | $\bar{g}$: nicht notwendig, ungewiß |

*Anmerkung.* Zu diesen Grundsymbolen werden später (außer dem Abkürzungszeichen $n$) noch zwei Wertsymbole $w$, $s$ mit ihren Überstrichenen hinzutreten, s. S. 83.

Die Zeichen $z, n = \bar{z}, m, \bar{m}, g, \bar{g}$ mögen als *Grundwerte* bezeichnet werden, und zwar die unüberstrichenen als positive, die überstrichenen (einschließlich $n$) als negative Grundwerte. Die Werte $m, \bar{m}, g, \bar{g}$ seien als *echte* Grundwerte bezeichnet, die Werte $z$ und $n = \bar{z}$ als unechte. Jede endliche Verknüpfungsreihe von Grundwerten heiße ein *Wert*, und zwar ein *positiver*, wenn die Anzahlen der in ihr auftretenden Überstreichungen und $n$-Zeichen zusammen eine gerade Zahl (oder 0) ergeben, ein *negativer* im anderen Falle.

Von den angegebenen Bedeutungen wird in den Beweisen kein Gebrauch gemacht. Zur Interpretation sind lediglich in der Einleitung und im Schlußteil einige Andeutungen gegeben; die inhaltliche Auswertung der Ergebnisse soll an anderer Stelle erfolgen.

*Fundamentale Axiome I:*

a) $zx = xz = x$ (für $x = z, \bar{z}, m, \bar{m}, g, \bar{g}$)

b) $nn = z$,

c) $\bar{m} = nm$ $\quad$ c⁰) $\bar{g} = ng$.

*Anmerkung* zum Prozeß des Schließens. Die Gleichheit genügt den üblichen Schlußregeln der Halbgruppe, die hier — im Hinblick auf die in Teil II

[4]) Dieses Axiom kommt nicht in idempotenten Modalitätenstrukturen der von Becker (4) vorgeschlagenen Implikation „was zutreffend ist, ist unmöglich unmöglich" gleich, wie schon Parry (8) zeigte — s. auch Churchman (7); man vgl. hierzu den Satz 26 des § 5.

[5]) Die neuere philosophische Diskussion über die Modalitäten unterscheidet ohnehin eine ganze Reihe sehr verschiedenartiger Möglichkeits- und Notwendigkeitsbegriffe.

erfolgende Erweiterung durch Schlußregeln für die Implikation — explizite angeführt seien:

*Schlußregeln der Gleichheit*, G ($x$ Grundwert, $\alpha$, $\beta$ beliebige Werte)

G 1) Transitivität: $\dfrac{\alpha = \beta \qquad \beta = \gamma}{\alpha = \gamma}$,

G 2) Symmetrie: $\dfrac{\alpha = \beta}{\beta = \alpha}$,

G 3) Rechtsmultiplikation: $\dfrac{\alpha = \beta}{\alpha x = \beta x}$ (gelesen: mit $\alpha = \beta$ ist auch $\alpha x = \beta x$)

G 4) Linksmultiplikation: $\dfrac{\alpha = \beta}{x\alpha = x\beta}$.

Definition der *Quaternalität.*

Zwei Werte mögen „zueinander dual" heißen, wenn sie durch *Vertauschung der Symbole $m$, $g$* auseinander hervorgehen. — Ist irgendein Wert durch das Symbol $\alpha$ mitgeteilt, so kann der dazu duale durch $\alpha^0$ mitgeteilt werden (z. B. für $\alpha = z\bar{m}ng$ ist $\alpha^0 = z\bar{g}nm$).

Zwei Werte mögen „zueinander verneinend" heißen, wenn sie auseinander dadurch hervorgehen, daß man im *ersten* auftretenden *echten* Grundwert Überstreichung mit Nichtüberstreichung vertauscht. (Falls kein echter Grundwert auftritt, ist die Vertauschung am ersten unechten vorzunehmen.) — Ist irgendein Wert durch das Symbol $\alpha$ mitgeteilt, so kann der dazu verneinende durch $\bar{\alpha}$ mitgeteilt werden (z. B. für $\alpha = z\bar{m}ng$ ist $\bar{\alpha} = zmng$).

*Satz 1.* $1_1$: $\alpha^{00} = \alpha$, $\bar{\bar{\alpha}} = \alpha$, $1_2$: $\bar{\alpha}^0 = \overline{\alpha^0}$, $1_3$: $(\alpha\beta)^0 = \alpha^0\beta^0$,
$1_4$: $\bar{\alpha} = n\alpha$, $1_5$: $\overline{\alpha\beta} = \bar{\alpha}\beta$.

$1_1$—$1_3$ sind symbolische Identitäten, die der Definition entfließen. $1_4$ sind Gleichheiten, die auf Grund der Axiome I gelten; $1_5$ folgt unmittelbar aus $1_4$.

Gemäß Satz $1_2$ erhält man zu jedem Wert (d. i. zu jeder Verknüpfungsreihe von Grundwerten) durch Verneinen und Dualisieren genau ein Quadrupel „*zueinander quaternaler*" Werte. — Beispiel:

1. $nm\bar{g}$, 2. $n\bar{m}\bar{g}$, 3. $ng\bar{m}$, 4. $n\bar{g}\bar{m}$. (Im folgenden wird stets so geordnet werden, daß 2 zu 1, 4 zu 3 verneinend, 3 zu 1, 4 zu 2 dual ist.)

Zwei *Gleichheiten* werden kurz verneinend bzw. dual bzw. quaternal genannt, wenn ihre linken und ebenso ihre rechten Seiten jeweils zueinander verneinend bzw. dual bzw. quaternal sind. So läßt sich zu jeder betrachteten Gleichheit ein (in sich) „*quaternales Quadrupel*" von Gleichheiten bilden (Beispiele folgen sogleich im nächsten Beweis).

*Satz 2.* Auf Grund der Axiome I und der Gleichheitsschlußregeln G sind mit einer Gleichheit zugleich die zu ihr quaternalen Gleichheiten beweisbar, m. a. W.: der Bestand der beweisbaren Gleichheiten ist „in sich quaternal".

*Beweis.*

1. Alle Gleichheiten des quaternalen Quadrupels

$\mathrm{Ia}_1$) $zm = mz\ (= m)$, $\mathrm{Ia}_2$) $z\bar{m} = \bar{m}z\ (= \bar{m})$,
$\mathrm{Ia}_3$) $zg = gz\ (= g)$, $\mathrm{Ia}_4$) $z\bar{g} = \bar{g}z = \bar{g}$

gelten nach Ax. Ia.

2. Von den Gleichheiten des quaternalen Quadrupels

$$\mathrm{Ic}_1)\ \overline{m} = nm, \quad \mathrm{Ic}_2)\ m = n\overline{m}, \quad \mathrm{Ic}_3)\ \overline{g} = ng, \quad \mathrm{Ic}_4)\ g = n\overline{g}$$

sind $\mathrm{Ic}_1$) und $\mathrm{Ic}_3$) die Axiome Ic bzw. $\mathrm{Ic}^0$. Die beiden anderen folgen unmittelbar, nämlich: $n\overline{m} = nnm$ (Ax. Ic) $= zm$ (Ib) $= m$ (Ia); entsprechend $ng = \overline{g}$.

3. Nach Abs. 1 und 2 sind die zu den Axiomen quaternalen Gleichheiten beweisbar. Es braucht daher lediglich noch gezeigt zu werden: Zu jedem nach den Schlußregeln G gültigen Schluß werden die quaternalen Schlüsse ableitbar, d. h.

Wenn $\varepsilon = \zeta$ aus $\gamma = \delta$ (bzw. aus $\gamma_1 = \delta_1$, $\gamma_2 = \delta_2$) folgt, so folgt: $3_1$. die duale Gleichheit $\varepsilon^0 = \zeta^0$ aus $\gamma^0 = \delta^0$ (bzw. aus $\gamma_1^0 = \delta_1^0$, $\gamma_2^0 = \delta_2^0$) und: $3_2$. die durch Verneinung entstehende Gleichheit $\overline{\varepsilon} = \overline{\zeta}$ aus $\overline{\gamma} = \overline{\delta}$ (bzw. aus $\overline{\gamma_1} = \overline{\delta_1}$, $\overline{\gamma_2} = \overline{\delta_2}$).

Beweis zu $3_1$. durch schrittweises Dualisieren der Beweisfigur, die von $\gamma = \delta$ zu $\varepsilon = \zeta$ führt:

Die Schlußregeln G 1, G 2 sind invariant gegen Dualisierung der Gleichheiten.

Zu G 3: Ein $\dfrac{\alpha = \beta}{\alpha x = \beta x}$ geht über in $\dfrac{\alpha^0 = \beta^0}{\alpha^0 x^0 = \beta^0 x^0}$, d. h. nach Satz $1_3$ und G 1 in $\dfrac{\alpha^0 = \beta^0}{(\alpha x)^0 = (\beta x)^0}$. — Zu G 4: entsprechend.

Beweis zu $3_2$. durch schrittweises Verneinen in der Beweisfigur, die von $\gamma = \delta$ zu $\varepsilon = \zeta$ führt:
G 1, G 2 sind invariant gegen Verneinung der Werte.

Zu G 3: Mit $\dfrac{\alpha = \beta}{\alpha x = \beta x}$ ist $\dfrac{\bar{\alpha} = \bar{\beta}}{\bar{\alpha} x = \bar{\beta} x}$, d. h. nach Satz $1_5$ und G 1 auch $\dfrac{\bar{\alpha} = \bar{\beta}}{\overline{\alpha x} = \overline{\beta x}}$ richtig.

Zu G 4: Mit $\dfrac{\alpha = \beta}{y\alpha = y\beta}$ ist $\dfrac{\alpha = \beta}{\bar{y}\alpha = \bar{y}\beta}$, d. h. nach Satz $1_5$ und G 1 auch $\dfrac{\alpha = \beta}{\overline{y\alpha} = \overline{y\beta}}$. Da außerdem $\dfrac{\bar{\alpha} = \bar{\beta}}{n\bar{\alpha} = n\bar{\beta}}$, d. h. nach Satz $1_1$ und $1_4$ $\dfrac{\bar{\alpha} = \bar{\beta}}{\alpha = \beta}$ ist, hat man $\dfrac{\bar{\alpha} = \bar{\beta}}{\overline{y\alpha} = \overline{y\beta}}$.

*Satz 3.* Bei Zufügung einer Gleichheit $\alpha = \beta$ zu den Axiomen I wird $\bar{\alpha} = \bar{\beta}$ beweisbar.

*Beweis.* Da auf beiden Seiten $n$ vorgesetzt werden darf (Schlußregel G 4), ist die Behauptung eine unmittelbare Folge von Satz $1_4$ und G 1, G 2.

Zur fundamentalen Axiomengruppe I tritt weiterhin in dieser Arbeit eine Reihe von „Kürzungsaxiomen" der Gestalt $x\,y = u$, — wo $x, y, u$ (i. allg.) Grundwerte bezeichnen. Es wird untersucht, inwieweit sich durch diese Axiome die Werte (d. s. die Verknüpfungsreihen von Grundwerten) auf bestimmte Zweierprodukte reduzieren.

Ein System von Sätzen — hier von Gleichheiten und später von Implikationen — die 1. von der jeweils vorausgeschickten Axiomenbasis unabhängig sind und 2. im Rahmen dieser Axiomenbasis einander äquivalent

sind, sei als „*Axiomablengruppe*" bezeichnet[6]). Es genügt, jeweils *einen* dieser Sätze als neues Axiom anzufügen, um die anderen beweisbar zu machen. Die Auswahl dieses Axioms ist willkürlich, d. h. hinsichtlich ihrer Axiomfähigkeit — als „*Axiomablen*" — sind die Sätze der Axiomablengruppe einander gleichberechtigt.

Die Angabe einer ganzen Axiomablengruppe gestattet eine systematische Übersicht über die jeweils in Frage kommenden Axiomenbasen und ihr gegenseitiges Verhältnis.

*Axiomablengruppe II* (Anhängung von $n$), bestehend aus einem Quadrupel quaternaler Gleichungen:

$$1.\ mn = \bar{g},\quad 2.\ \bar{m}n = g,\quad 3.\ gn = \bar{m},\quad 4.\ \bar{g}n = m.$$

*Satz 4.* Die vier Gleichheiten II sind einander im Rahmen der Axiomenbasis I äquivalent; es genügt also, irgend*eines* von ihnen als Axiom anzufügen.

*Beweis.*

Erstens: Aus $\alpha = \beta$ folgt $\bar{\alpha} = \bar{\beta}$ nach Satz 3. Daher folgt 2. aus 1., 1. aus 2., 4. aus 3., 3. aus 4.

Zweitens: Aus 1. folgt 4.

$\bar{g}n = mnn$ (Gleichung 1 und G 2, 3) $= mz$ (I b, G 4) $= m$ (I a).

Drittens: Aus 4. folgt 1.

$mn = \bar{g}nn$ (Gleichung 4 und G 2, 3) $= \bar{g}z$ (I b, G. 4) $= \bar{g}$ (I a).

*Satz 5.* Auf Grund der Axiome I, eines der Axiomablen II und der Schlußregeln für die Gleichheit, G, gilt

$$\alpha n = n\alpha^0 = \overline{\alpha^0}.$$

*Beweis* durch vollständige Induktion (wobei $\bar{\beta} = n\beta$ nach Satz $1_4$ benutzt wird). Für einen Grundwert $x$ gilt $xn = \overline{x^0}$, nämlich für ein unechtes $x$ nach den Ax. I, für echtes $x$ nach den Ax. II, die nach Satz 4 einander äquivalent sind. — Sei nun $\alpha = \beta\gamma$.

$\alpha n = \beta\gamma n = \beta n \gamma^0$ (Induktionsannahme) $= \overline{\beta^0}\gamma^0$ (Induktionsannahme) $=$
$= \overline{\beta^0\gamma^0}$ (Satz $1_5$) $= \overline{(\beta\gamma)^0}$ (Satz $1_3$).

*Satz 6.* Bei Zufügung einer Gleichheit $\alpha = \beta$ zu den Axiomen I und einem der Axiomablen II wird $\alpha^0 = \beta^0$ beweisbar.

*Beweis.* Mit $\alpha = \beta$ ist $\alpha n = \beta n$ nach G 3, also $n\alpha^0 = n\beta^0$ nach Satz 5 und G 1, 2, also $nn\alpha^0 = nn\beta^0$ nach G 4 und endlich $\alpha^0 = \beta^0$ nach Ax. I b und G 1–3.

*Satz 7. Quaternalitätssatz* für Gleichheiten. Im Rahmen irgendeines Axiomensystems, das die fundamentalen Axiome I und eines der Axiomablen II enthält, sind (bei Benutzung der Gleichheits-Schlußregeln G) mit einer Gleichheit zugleich die zu ihr quaternalen Gleichheiten beweisbar. M. a. W.: der Bestand der beweisbaren Gleichheiten ist „in sich quaternal".

*Beweis.*

Auf Grund der Axiome I und eines der Axiomablen II sind bei Hinzutreten irgendeiner Gleichheit die quaternalen Gleichheiten nach Satz 3 und Satz 6 beweisbar.

[6]) Der Terminus schließt sich an den Terminus „Axiomengruppe" an; es handelt sich natürlich nicht um eine „Gruppe" im algebraischen Sinne.

*Satz 8.* Auf Grund der Axiome I, eines der Axiomablen II und der Schlußregeln G gelten die folgenden beiden quaternalen Quadrupel von Gleichheiten zwischen Zweierprodukten echter Grundwerte („Kommutatorformeln"):

(P) 1. $mm = \bar{g}\bar{m}$, 2. $\bar{m}m = g\bar{m}$, 3. $gg = \bar{m}\bar{g}$, 4. $\bar{g}g = m\bar{g}$,

(Q) 1. $m\bar{m} = \bar{g}m$, 2. $\bar{m}\bar{m} = gm$, 3. $g\bar{g} = \bar{m}g$, 4. $\bar{g}\bar{g} = mg$.

*Beweis:* Für (P) 1.: $mm = mn\bar{m}$ ($\dot{\mathrm{I}}c_2$) $= \bar{g}\bar{m}$ (II 1).

Für (Q) 1.: $m\bar{m} = mnm$ ($\mathrm{I}c_1$) $= \bar{g}m$ (II 1).

Dazu der Quaternalitätssatz 7.

*Axiomablengruppe III*, bestehend aus zwei quaternalen Quadrupeln:

a 1. $mm = m$, a 2. $\bar{m}m = \bar{m}$, a 3. $gg = g$, a 4. $\bar{g}g = \bar{g}$,

b 1. $g\bar{m} = \bar{m}$, b 2. $\bar{g}\bar{m} = m$, b 3. $m\bar{g} = \bar{g}$, b 4. $\bar{m}\bar{g} = g$.

*Satz 9.* Die Axiomablen III sind einander im Rahmen der Axiome I, II (und der Schlußregeln G) äquivalent; es genügt also auch hier, *eines* von ihnen als Axiom zuzufügen.

*Beweis.* Innerhalb der Quadrupel: Quaternalitätssatz 7. Dazu:

IIIa 1 $\rightarrow$ IIIb 3: $m\bar{g} = mmn$ (II 1) $= mn$ (IIIa 1) $= \bar{g}$ (II 1);

IIIb 3 $\rightarrow$ IIIa 4: $\bar{g}g = mng$ (II 1) $= m\bar{g}$ ($\mathrm{I}\,c^0$) $= \bar{g}$ (IIIb 3).

## § 3. Modalitätenstrukturen (Definitionen; allgemeine Eigenschaften).

Definition (M) der *Modalitätenstruktur* und ihrer Stufe.

Ein Bereich von Gleichheiten zwischen den auf S. 73 (im vorletzten Absatz) eingeführten Werten möge eine Modalitätenstruktur heißen, wenn

1. die Axiomablen I, II zu ihm gehören,
2. die Schlußregeln G nicht aus dem Bereich hinausführen,
3. kein positiver Wert (S. 73) einem negativen Wert gleich ist,
4. $m$ nicht gleich $g$ ist.

Eine solche Modalitätenstruktur möge von *p-ter Stufe* heißen, wenn in ihr

1. jeder Wert einem solchen gleich ist, der sich aus höchstens $p$ Grundwerten zusammensetzt,

2. ein aus $p$ Grundwerten zusammengesetzter Wert vorkommt, der keinem kürzeren gleich ist.

Eine Modalitätenstruktur heißt *idempotent*, wenn in ihr $mm = m$ (IIIa 1) und mithin nach Satz 9 auch $gg = g$ (IIIa 3) gelten.

*Satz 10.* Ein endlicher Bereich von Gleichheiten, der die Axiome I und ein Axiomabel II umfaßt und die Bedingung (M 3) erfüllt, läßt sich stets als *Axiomensystem* einer Modalitätenstruktur auffassen, die mittels der Regeln G hervorgeht, sobald nur gezeigt ist, daß $m$ nicht gleich $g$ werden kann.

*Beweis.* (M 3) ist für den hervorgehenden Bereich von Gleichheiten erfüllt, denn die Regeln G führen von Gleichheiten, die (M 3) erfüllen, d. h. von Gleichheiten zwischen positiven oder zwischen zwei negativen Werten, stets auf Gleichheiten derselben Art.

Definition der *Wertebasis*. Unter einer Wertebasis einer Modalitätenstruktur sei eine Menge paarweise verschiedener, d. h. hier: nicht beweisbar gleicher Werte derart verstanden, daß jeder beliebige Wert der Struktur einem dieser „Basiswerte" gleich ist.

*Anmerkung*. Gemäß dieser Definition gibt es zu einer Modalitätenstruktur „bis auf Gleichheit" genau eine Wertebasis, genauer: zwei Wertebasen sind umkehrbar eindeutig aufeinander abbildbar, wobei zugeordnete Werte einander gleich sind.

*Satz 11*. Die Wertebasis einer Modalitätenstruktur enthält alle sechs Grundwerte (genauer: jede Wertebasis enthält zu jedem Grundwert einen gleichen Wert).

*Folgerungen*. 1. Ein Bereich von Gleichheiten, der eine Gleichung zwischen nichtidentischen Grundwerten enthält, ist keine Modalitätenstruktur. 2. „Die" Wertebasis einer einstufigen Modalitätenstruktur wird von den sechs Grundwerten gebildet.

Zum *Beweis* des Satzes 11 genügt es zu zeigen, daß die Grundwerte in einer beliebigen Modalitätenstruktur voneinander verschieden sind.

Nach (M 3) und (M 4) bleibt nur nachzuweisen, daß keine der Gleichheiten $z = m$, $z = g$, $n = \overline{m}$, $n = \overline{g}$, $\overline{m} = \overline{g}$ ableitbar ist. Nach dem Quaternalitätssatz 7 folgen die ersten vier dieser Gleichungen aus einer jeden von ihnen; jede führt somit mit G 1, 2 auf $g = m$, entgegen (M 4). — $\overline{m} = \overline{g}$ führt nach dem Quaternalitätssatz mit G 2 unmittelbar ebenfalls auf $g = m$.

*Satz 12*. —$12_1$. Die Axiome[7]) I—III, d. h. nach den Sätzen 4 und 9 zugleich: die Axiome I, irgendeines der Axiomablen II und irgendeines der Axiomablen III spannen mit den Regeln G zusammen bereits eine idempotente Modalitätenstruktur auf. —$12_2$. Diese ist nicht von endlicher Stufe, d. h. auf Grund der genannten Axiome und Regeln reduzieren sich die Werte (d. s. die Verknüpfungsreihen von Grundwerten) nicht auf endlich viele.

Dem Beweis seien eine Definition und einige Hilfssätze vorangeschickt, die auch weiterhin noch Verwendung finden werden.

Rekursive Definition der „*m-n-Reduzierten*" $R$ eines beliebigen Wertes: $Rz$ leer (die leere Verknüpfungsreihe), $Rn = n$, $Rm = m$, $R\overline{m} = nm$, $Rg = nmn$, $R\overline{g} = mn$; $R\alpha x = R\alpha \cdot Rx$ für einen beliebigen Wert $\alpha$ und einen Grundwert $x$.

*1. Hilfssatz*. Ist $\alpha = \beta$ auf Grund der Axiome I—III und der Schlußregeln G beweisbar, so ist in einer Halbgruppe mit den Elementen $m$ und $n$ und den definierenden Relationen

$$\text{R 1.} \quad nn = 1, \qquad \text{R 2.} \quad mm = m,$$

das „Wort" $R\alpha$ dem Wort $R\beta$ gleich. (Statt des Ausdrucks „$m$-$n$-Wert" wird bei Überlegungen zu den Halbgruppen auch der Ausdruck „Wort" gebraucht.)

Zum *Beweis* genügt es, rechnerisch zu bestätigen:

1. Auf Grund der Axiome I, II ist für jeden beliebigen Wert $\alpha = R\alpha$.

2. Daher trifft die Behauptung auf die Axiome I, II, III selbst zu,

---

[7]) Sobald ein Axiomabel in ein Axiomensystem aufgenommen ist, fungiert es als Axiom.

3. und wenn eine Schlußregel G von $\gamma = \delta$ auf $\varepsilon = \zeta$ führt, so führt $R\gamma = R\delta$ auf $R\varepsilon = R\zeta$ (analog für die Schlußregeln mit *zwei* Voraussetzungsgleichheiten). — Die Schlußregeln G sind aber sämtlich für eine Halbgruppe gültig.

Definitionen zum Bau der in Hilfssatz 1 definierten $m$-$n$-Halbgruppe.

Zwei Zeichen eines „Wortes" sollen derselben „*$n$-Abteilung*" angehören, wenn zwischen ihnen kein $m$ steht. Eine $n$-Abteilung aus ungerade vielen $n$-Zeichen heiße *echt*. — Zwei $m$-Zeichen eines „Wortes" sollen derselben *$m$-Abteilung* angehören, wenn keine echte $n$-Abteilung zwischen ihnen steht. — Eine $n$-Abteilung, auf die kein $m$ mehr folgt, heiße „abschließend"; entsprechend: eine $m$-Abteilung, der kein $n$ vorangeht, heiße „einleitend" (beide brauchen offenbar in einem gegebenen Wort nicht zu existieren).

2. *Hilfssatz.* Das $2p$-gliedrige Wort $(nm)^p$ — bei natürlichem $p$ — ist in der in Hilfssatz 1 genannten Halbgruppe nicht auf ein kürzeres Wort reduzierbar.

Zum *Beweis* genügt es einzusehen, daß durch Umformungen auf Grund der definierenden Relationen R 1., R 2. kein Wort die Zahl seiner echten $n$-Abteilungen oder die Zahl seiner $m$-Abteilungen ändert.

*Beweis* des Satzes 12. — Die in Satz 12 genannten Axiome I—III erfüllen selbst die Bedingung (M 3); nach Satz 10 erfüllen also alle mittels der Regeln G folgenden Gleichheiten diese Bedingung. Nach dem 1. Hilfssatz wäre bei $m = g$ in der Halbgruppe die Gleichheit $m = nmn$ beweisbar, entgegen der im Beweise des 2. Hilfssatzes ausgesprochenen Behauptung. Daher ist auch die Bedingung (M 4) erfüllt. Hiermit ist Satz $12_1$ bewiesen.

Da die $m$-$n$-Reduzierten der Grundwerte höchstens ein $m$-Zeichen enthalten, läßt sich nach dem 1. und 2. Hilfssatz der Wert $\overline{m}^p$ (mit natürlichem $p$) nicht auf eine Verknüpfung von weniger als $p$ Grundwerten reduzieren. Es gilt also auch $12_2$.

*Satz 13.* Isomorphe idempotente Modalitätenstrukturen stimmen überein.

*Beweis.* Die Abbildung eines Wertes $\alpha$ auf den Wert $\beta$ bei isomorphen Modalitätenstrukturen sei durch $\alpha \equiv \beta$ abgekürzt.

1. In isomorphen Modalitätenstrukturen ist $z \equiv z$:

Nach Ax. II 1 ist $mn = \overline{g}$. Wäre $m \equiv z$ oder $n \equiv z$, so müßten nach Ax. I a $n$ und $\overline{g}$ bzw. $m$ und $\overline{g}$ auf ein und denselben Wert abgebildet sein; die Abbildung wäre also nicht umkehrbar eindeutig. Daher ist $m \not\equiv z$, $n \not\equiv z$. Dual erhält man aus Ax. II 3: $g \not\equiv z$.

Entsprechend schließt man aus Ax. II 2 ($\overline{m}n = g$) auf $\overline{m} \not\equiv z$ und dual aus Ax. II 4 auf $\overline{g} \not\equiv z$. Da $z$ das Bild eines Grundwertes sein muß, bleibt nur $z \equiv z$.

2. In isomorphen idempotenten Modalitätenstrukturen ist $n \equiv n$:

Nach Ax. III a 1 ist $mm = m$. Wäre $m \equiv n$, so müßte nach Ax. I b auch $m \equiv z$ sein; also $m \not\equiv n$. Dual erhält man aus Ax. III a 3: $g \not\equiv n$.

Nach (Q 4) ist $\overline{g}\overline{g} = mg$. Wäre $\overline{g} \equiv n$, so müßte nach Ax. I b auch $mg \equiv z$ sein. Da nun bereits $z \equiv z$ bewiesen ist, müßte $mg = z$ sein. Dies würde aber $z = mg = mmg$ (III 1a) $= mz = m$ (I a) nach sich ziehen, was nach der 1. Folgerung des Satzes 11 nicht sein kann. Mithin ist $\overline{g} \not\equiv n$. Dual erhält man $\overline{m} \not\equiv n$. — Da $n$ das Bild eines Grundwertes sein muß, ist $n \equiv n$.

3. In isomorphen idempotenten Modalitätenstrukturen ist $m \equiv m$ oder $m \equiv g$:

Nach Ax. IIIa 1 ist $mm = m$. Aus $m \equiv \bar{m}$ würde folgen $\bar{m}\bar{m} = \bar{m}$, aus $m \equiv \bar{g}$ würde folgen $\bar{g}\bar{g} = \bar{g}$. Beides steht im Widerspruch zu (M 3). Daher ist $m \not\equiv \bar{m}$, $m \not\equiv \bar{g}$. Da $m$ auf einen Grundwert abgebildet sein muß, bleibt nach den Absätzen 1, 2 dieses Beweises nur: $m \equiv m$ oder $m \equiv g$.

4. Bei $m \equiv m$ hat man die *identische* Abbildung:

Mit $z \equiv z$, $n \equiv n$, $m \equiv m$ hat man nach Ax. Ic: $\bar{m} \equiv \bar{m}$ und mithin nach Ax. II 2: $g \equiv g$ und weiter nach Ax. Ic$^0$: $\bar{g} \equiv \bar{g}$.

5. Bei $m \equiv g$ hat man die *duale* Abbildung:

Mit $z \equiv z$, $n \equiv n$, $m \equiv g$ hat man nach Ax. Ic und Ic$^0$: $\bar{m} \equiv \bar{g}$ und mithin nach Ax. II 2 und II 4: $g \equiv m$, daraus weiter nach Ax. Ic$^0$ und Ic: $\bar{g} \equiv \bar{m}$. Dazu Satz $1_3$.

6. Nach Abs. 3, 4, 5 dieses Beweises bleibt als einzige nicht identische Isomorphie die Dualität. Der Quaternalitätssatz 7 lehrt, daß diese duale Abbildung einerseits für jede Modalitätenstruktur möglich ist, andererseits jedoch sie in sich selbst überführt.

*Satz 14.* — $14_1$. Zu den Axiomen I—III — genauer (gemäß den Sätzen 4 und 9): zu einem Axiomensystem, das aus den Axiomen I und je einem Axiomabel II, III besteht — braucht man, um (mit den Regeln G) eine idempotente Modalitätenstruktur von *beliebig* gegebener Stufe aufzuspannen, nur ein einziges Axiom zuzufügen. — $14_2$. Für eine Stufe $p + 1 > 1$ genügt dazu eine der beiden (nach Satz 7 äquivalenten) Gleichheiten:

*Axiomablengruppe* $\Sigma$. 1. $\bar{m}^{p+2} = \bar{m}^p$; $\bar{g}^{p+2} = \bar{g}^p$ $(p > 0)$.

*Anmerkung*: Eine *ein*stufige idempotente Modalitätenstruktur der in $14_1$ verlangten Art wird im nächsten Paragraphen ohne Benutzung des allgemeinen Satzes $14_1$ in concreto vorgeführt werden.

Beweis des Satzes 14 für eine Stufe $p + 1 > 1$. (Nach Satz 7 kann dabei das Axiomabel $\Sigma\,1$ zugrundegelegt werden.)

a) Wie im 1. Hilfsatz zu Satz 12 sieht man ein: Ist $\alpha = \beta$ im neuen Axiomensystem beweisbar, so ist in einer Halbgruppe mit den Elementen $m$ und $n$ und den definierenden Relationen

$$\text{R1. } nn = 1, \quad \text{R2. } mm = m, \quad \text{R3. } (nm)^{p+2} = (nm)^p$$

das Wort $R\,\alpha$ dem Wort $R\,\beta$ gleich, und umgekehrt.

Jede $m$-$n$-Reduzierte eines Wertes, $R\,\alpha$, läßt sich auf Grund der Relationen R 1., R 2. in ein solches „Wort" überführen, in dem die $m$- und $n$-Zeichen alternieren, und somit weiter gemäß R 3. in ein Wort, das eine der Gestalten hat:

(1) leer, $n$, $(nm)^q$, $(nm)^q \cdot n$, $n\,(nm)^q$, $n\,(nm)^q \cdot n$ — mit $1 \leq q \leq p + 1$.

Dann aber geht $\alpha$ (mit Benutzung von Abs. 1 des Beweises zum Hilfssatz auf S. 78 unten) in eine der Gestalten:

$$\text{leer, } n,\ \bar{m}^q,\ \bar{m}^q \cdot n = \bar{m}^{q-1} \cdot g,\ n \cdot \bar{m}^q = m \cdot \bar{m}^{q-1},\ n\bar{m}^q \cdot n \begin{cases} = n\bar{m}^{q-2} \cdot g & \text{bei } q > 1 \\ = g & \text{bei } q = 1 \end{cases}$$

über. Die Modalitätenstruktur ist also von höchstens $p + 1$-ter Stufe. —

b) Ein Wort mit $r \leqq p - 1$ $m$-Abteilungen (S. 79) kann durch Anwendung von R 1. und R 2. allein nur wieder in ein Wort mit $r$ $m$-Abteilungen übergeführt werden; auf ein solches ist R 3. nicht anwendbar. Daher kann ein Wort mit $r \leqq p - 1$ $m$-Abteilungen überhaupt nur in ein Wort mit $r$ $m$-Abteilungen übergeführt werden. — c) Irgendein Wort mit ungerade vielen $m$-Abteilungen geht durch Umformungen gemäß R 1., R 2., R 3. stets nur in wieder in ein Wort mit ungerade vielen $m$-Abteilungen über. — d) Das Wort $(nm)^{p+1}$ hat $p + 1$ $m$-Abteilungen. Würde es sich in ein Wort mit weniger $m$-Abteilungen, d. h. nach c) mit $r \leqq p - 1$ $m$-Abteilungen überführen lassen, so hätte man einen Widerspruch zu b). Daraus folgt: $\overline{m}^{p+1}$ ist nicht einem kürzeren Produkt aus Grundwerten gleich.

e) Eine echte einleitende oder abschließende $n$-Abteilung kann nicht durch Umformung beseitigt werden.

f) Die in Satz 14 aufgeführten Axiome I—III und $\Sigma$ erfüllen die Bedingung (M 3). Daher erfüllt nach dem Beweise für Satz 10 der mit den Regeln G aufgespannte Bereich ebenfalls diese Bedingung. — Der zu Anfang des Abs. a) eingeführte Hilfsatz lehrt, daß bei $m = g$ in der Halbgruppe die Gleichheit $m = nmn$ erfüllt wäre, entgegen e); mithin ist auch (M 4) erfüllt. Man hat also im System des Satzes $14_2$ eine Modalitätenstruktur vor sich. Gemäß Abs. d) ist diese mindestens von $p + 1$-ter Stufe, also nach Abs. a) genau von $p + 1$-Stufe. — Hiermit ist Satz 14 für $p + 1 > 1$ bewiesen.

*Satz 15.* Die in Satz 14 angeführte $q(= p + 1)$-stufige idempotente Modalitätenstruktur hat eine aus $4q + 2$ Werten bestehende Basis.

Zum Beweis: Gemäß Abs. a) des Beweises für Satz 14 ist jeder Wert einem der Werte (1) gleich; die Absätze b), c), e) jenes Beweises lehren außerdem, daß diese Werte voneinander verschieden sind.

## § 4. Die einstufigen idempotenten Modalitätenstrukturen.

*Hilfssatz.* Jede einstufige idempotente Modalitätenstruktur läßt sich durch Zufügung *eines* Axioms zur Axiomenbasis I—III (mit den Regeln G) erhalten; und zwar werden alle einstufigen idempotenten Modalitätenstrukturen insbesondere durch Angabe eines Grundwertes, der dem Produkt $mg$ gleich sein soll, aufgespannt.

*Beweis.* Die Axiomenbasis I—III läßt von den 36 Zweierprodukten, die sich aus Grundwerten bilden lassen, nur die acht in den Gleichungen (Q) des Satzes 8 auftretenden Zweierprodukte unreduziert. Da diese sich zu dem quaternalen Gleichheitsquadrupel (Q) ordnen, so genügt nach dem Quaternalitätssatz 7 die Reduktion *eines* solchen Zweierproduktes auf einen Grundwert, um jedes Zweierprodukt und mithin (durch Induktion) jedes Produkt, d. h. jeden Wert auf einen Grundwert zu reduzieren; man braucht z. B. nur für $mg$ eine solche Reduktion anzugeben.

Eine weitere Gleichung $\alpha = \beta$, die nicht beweisbar wird, läßt sich nicht mehr axiomatisch zufügen, denn: sei $\alpha$ auf den Grundwert $x$, $\beta$ auf den Grundwert $y$ reduziert. Falls $x$ und $y$ identisch sind, ist $\alpha = \beta$ beweisbar, falls nicht, so würde aus $\alpha = \beta$ auch $x = y$ folgen, was nach der 1. Folgerung zu Satz 11 verboten ist. Daher ist die einstufige Modalitätenstruktur durch Angabe eines Grundwertes für $mg$ in unerweiterbarer Weise festgelegt; hiermit ist der Hilfsatz bewiesen.

Nach einem Schluß aus Abs. 2 des Beweises für Satz 13 ist $mg = z$ verboten. Es bleiben also höchstens zwei einstufige Modalitätenstrukturen, die sich durch die Forderung $mg = m$ bzw. $mg = g$ charakterisieren lassen. Die beiden zugehörigen, konkurrierenden Axiomablengruppen IV, V entsprechen einander dual in den *rechten* Seiten der betreffenden Gleichungen; jede umfaßt zwei quaternale Quadrupel.

| *Axiomablengruppe IV.* | | *Axiomablengruppe V.* | |
|---|---|---|---|
| a1. $m\overline{m} = \overline{m}$, | a2. $\overline{m}\overline{m} = m$. | a1. $m\overline{m} = \overline{g}$, | a2. $\overline{m}\overline{m} = g$, |
| a3. $g\overline{g} = \overline{g}$, | a4. $\overline{g}\overline{g} = g$, | a3. $g\overline{g} = \overline{m}$, | a4. $\overline{g}\overline{g} = m$, |
| b1. $gm = m$, | b2. $\overline{g}m = \overline{m}$, | b1. $gm = g$, | b2. $\overline{g}m = \overline{g}$, |
| b3. $mg = g$, | b4. $\overline{m}g = \overline{g}$, | b3. $mg = m$, | b4. $\overline{m}g = \overline{m}$. |

(Bemerkungen zur Interpretation werden demnächst an anderer Stelle folgen.)

*Anmerkung.* Das System III, IV und ebenso das System III, V geben für alle Zweierprodukte echter Grundwerte einen Grundwert an; mit I, II zusammen hat man für jedes Zweierprodukt von Grundwerten einen Grundwert.

*Satz 16.* Im Rahmen der Axiome I, II allein — ohne III — sind alle Axiomablen der Gruppe IV und ebenso diejenigen der Gruppe V einander äquivalent; es genügt also, *eine* Gleichung IV oder *eine* Gleichung V als neues Axiom anzufügen.

*Beweis.* Innerhalb der Quadrupel: Quaternalitätssatz 7. Dazu: IVa 3 $\leftrightarrow$ IVb 4 und ebenso Va 3 $\leftrightarrow$ Vb 4 nach (Q) 3 (S. 77).

*Satz 17.* Aus jedem Axiomabel der Gruppe IV und ebenso aus jedem der Gruppe V folgen — im Rahmen der Axiome I, II — alle Axiomablen der Gruppe III.

*Beweis.* Nach den Sätzen 9 und 16 genügt es zu zeigen: IIIa 1 folgt aus IV und ebenso aus V.

$$mm = m\overline{m}\overline{m} \text{ (IVa 2)} = \overline{m}\overline{m} \text{ (IVa 1)} = m \text{ (IVa 2)}.$$
$$mm = \overline{g}\overline{g}\cdot m \text{ (Va 4)} = \overline{g}\overline{g} \text{ (Vb 2)} = m \text{ (Va 4)}.$$

*Satz 18.* Jede der Axiomenbasen I, II, IV und I, II, V (mit Regeln G) spannt eine idempotente Modalitätenstruktur auf, d. h. (gemäß den Sätzen 4, 9, 16) ausführlich: Die Gleichheiten, die aus den Axiomen I, einem Axiomabel II und einem Axiomabel IV oder V beweisbar werden, erfüllen die Definition (M) und umfassen die Axiomablen der Idempotenz IIIa 1, 3.

*Bezeichnung.* Die von I, II, IV aufgespannte Modalitätenstruktur möge die „*erste Sechsmodalitätenstruktur*" heißen, die von I, II, V aufgespannte die „*zweite Sechsmodalitätenstruktur*" (in beiden Fällen besteht gemäß der 2. Folgerung zu Satz 11 die Wertebasis aus den sechs Werten $z$, $m$, $g$, $n$, $\overline{m}$, $\overline{g}$).

Beweis des Satzes 18. Erster Teil: zum System I, II, IV.

Wie im Beweise zu Satz 12 erhält man: ist $\alpha = \beta$ auf Grund der Axiome I, II, IV und der Schlußregeln G beweisbar, so ist in einer Halbgruppe mit den Elementen $m$ und $n$ und den definierenden Relationen

$$\text{R } 1'.\ nn = 1 \text{ und R } 2'.\ nmnm = m \text{ (analog zu IVa 2)}$$

das „Wort" $R\alpha$ dem Wort $R\beta$ (Def. S. 78) gleich, und umgekehrt.

(Anmerkung. Satz 17 zeigt übrigens in diesem Zusammenhang, daß in einer Halbgruppe mit den definierenden Relationen R 1'., R 2'. das Produkt $mm$ auf $m$ reduzierbar ist, was sich auch unschwer direkt einsehen läßt).

Man ersieht nun:

1. bei keiner der Umformungen, die auf Grund der definierenden Relationen gestattet sind, ändert sich die Gerad- bzw. Ungeradzahligkeit der Anzahl der echten $n$-Abteilungen,

2. bei keiner dieser Umformungen ändert sich die Echtheit bzw. Unechtheit der abschließenden $n$-Abteilung.

3. keine der Umformungen gestattet, in ein Wort, das das $m$-Zeichen nicht enthält, ein solches hineinzubringen.

Diese drei Tatsachen lehren: die $m$-$n$-Reduzierten der Grundwerte lassen sich paarweise nicht ineinander überführen, — insbesondere ist also (M 4) erfüllt.

(M 3) ist nach Satz 10 ebenfalls erfüllt.

Zweiter Teil des Beweises: zum System I, II, V.

Man hat hier die definierenden Relationen

$$\text{R } 1''.\ nn = 1,\quad \text{R } 2''.\ mnmn = m \quad \text{(analog zu V b 3)}.$$

Der Beweis verläuft ganz entsprechend wie im ersten Teile; in Abs. 2 ist statt der abschließenden $n$-Abteilung die einleitende heranzuziehen.

*Satz 19.*—$19_1$. Die von den Axiomenbasen I, II, IV bzw. I, II, V (mit den Regeln G) aufgespannten Modalitätenstrukturen — die „erste" bzw. „zweite Sechsmodalitätenstruktur" — sind die einzigen einstufigen idempotenten Modalitätenstrukturen. — $19_2$. Diese beiden Modalitätenstrukturen sind nicht isomorph.

*Beweis.* — $19_1$ folgt aus den Sätzen 18 und 16 mit dem 1. Absatz von S. 82.

Wenn die beiden Strukturen übereinstimmen würden, so hätte man aus IV a 2 und V a 2: $m = g$, im Widerspruch zu (M 4). Nun folgt $19_2$ mit Satz 13.

## § 5. Die zweistufigen idempotenten Modalitätenstrukturen.

Die Axiomenbasis I—III (mit Schlußregeln G) läßt, wie im vorigen Paragraphen bereits erwähnt wurde, lediglich diejenigen acht Zweierprodukte von Grundwerten offen, die durch die Gleichheiten (Q) des Satzes 8 miteinander verbunden sind. Jede überhaupt erlaubte Gleichsetzung eines dieser Werte mit einem Grundwert führt (wie im Beweis des Hilfssatzes, der den § 4 einleitet, gezeigt wurde) auf eine einstufige Struktur. Daher bleiben alle diese Produkte in einer mehrstufigen idempotenten Modalitätenstruktur ungekürzt. Es erscheint in mancher Hinsicht zweckmäßig, kurze Symbole und Bezeichnungen für sie einzuführen:

*Abkürzungen*:

D 1. $w = \overline{m}\overline{m} = gm$ (Q 2) D 2. $s = \overline{g}\overline{g} = mg$ (Q 4),

D 3. $\overline{w} = m\overline{m} = \overline{g}m$ (Q 1) D 4. $\overline{s} = g\overline{g} = \overline{m}g$ (Q 3).

$w$, d. i. unmöglich unmöglich = gewiß möglich (notwendig möglich) sei kurz als „*widerspruchsfrei*" (unwiderlegbar), $s$, d. i. möglicherweise gewiß (möglicherweise notwendig) als „*stringenzfähig*" bezeichnet.

$w$ und $s$ sind definitionsgemäß *dual zueinander*; daher erkennt man mit Benutzung des Satzes $1_3$ unmittelbar:

*Satz 20.* — $20_1$. Auf Grund der Definitionen D und der Axiome I, II (mit Regeln G) gelten die folgenden Gleichheiten samt den ihnen durch Verneinung zugeordneten:

(S) $w = \bar{s}n,\ s = \bar{w}n;\ wm = \bar{s}\bar{m},\ sg = \bar{w}\bar{g};\ gw = \bar{m}\bar{w},\ ms = \bar{g}\bar{s};$

(T) $mw = mgm = sm\ (= \bar{g}\bar{w} = \bar{w}\bar{m}),\ gs = gmg = wg\ (= \bar{m}\bar{s} = \bar{s}\bar{g}).$

$20_2$. Auf Grund der Definitionen D und der Axiome I—III (mit Regeln G) gilt

(U) $wm = w = gw$ (nämlich $wm = gmm = gm = ggm = gw$), dual $sg = s = ms$.

(Anmerkung. Die systematische Ersetzung jedes $w$, $s$ durch die beiden es definierenden Zweierprodukte liefert übrigens die Formeln (P), (Q) des Satzes 8 zurück.)

*Hilfssatz.* Jede zweistufige idempotente Modalitätenstruktur läßt sich durch Zufügung *eines* Axioms zur Basis I—III (genauer: zu einer Basis, die aus den Axiomen I und je einem Axiomabel II, III besteht) — sowie eventuell der weiteren Gleichheit $w = s$ — mit den Regeln G erhalten, und zwar werden alle zweistufigen idempotenten Modalitätenstrukturen insbesondere z. B. durch Angabe eines Grundwertes oder Produktes zweier Grundwerte, der bzw. das dem Dreierprodukt $mw$ ($= m\bar{m}\bar{m}$) gleich sein soll, erzeugt.

*Beweis.* Man erkennt bei Durchsicht aller Axiomabeln I—III und der Gleichheiten (S), (U), daß lediglich diejenigen Zweierprodukte aus einem Grundwert und einem der Werte $w$, $\bar{w}$, $s$, $\bar{s}$ — und somit auch lediglich diejenigen Dreierprodukte aus Grundwerten nicht auf höchstens zwei Grundwerte reduzierbar sind, die durch die Gleichheiten (T) oder durch die bei Verneinung daraus entstehenden Gleichungen verbunden sind.

Jede zweistufige idempotente Modalitätenstruktur muß insbesondere für $mw$ eine Reduktion auf ein Zweierprodukt aus bloßen Grundwerten angeben; gemäß den Gleichungen (T) und dem Quaternalitätssatz 7 hat man dann für *jedes* Zweierprodukt aus einem Grundwert und einem der Werte $w$, $\bar{w}$, $s$, $\bar{s}$ eine Reduktion auf ein Zweierprodukt aus Grundwerten; somit hat man für jedes Dreierprodukt und überhaupt für jedes Produkt eine Reduktion auf einen Grundwert oder ein Produkt von zwei Grundwerten.

Es ist zu fragen, ob sich noch eine weitere nicht beweisbare Gleichung $\alpha = \beta$ als Axiom zufügen läßt. Man hat bereits eine Reduktion $\alpha = x$ (wo $x$ Grundwert ist) oder $\alpha = w, \bar{w}, s$ oder $\bar{s}$, entsprechend $\beta = y$ (wo $y$ Grundwert ist) oder $\beta = w, \bar{w}, s$ oder $\bar{s}$. — Bei $\alpha = x$, $\beta = y$ würde $x = y$ folgen; diese Gleichung ist nach der 1. Folgerung des Satzes 11 entweder schon beweisbar oder widerspruchsvoll. — Bei $\alpha = x$, $\beta = w, \bar{w}, s$ oder $\bar{s}$ hat man $x = w, \bar{w}, s$ oder $\bar{s}$; die Struktur wird dann einstufig (wie man mit dem Quaternalitätssatz erkennt). Der Fall $\alpha = w, \bar{w}, s$ oder $\bar{s}$, $\beta = y$ ist entsprechend. — Es bleibt der Fall $\alpha = w, \bar{w}, s$ oder $\bar{s}$, $\beta = w, \bar{w}, s$ oder $\bar{s}$. Wenn beide Male rechts derselbe Wert steht, ist $\alpha = \beta$ beweisbar. Sonst bleibt nach (M 3) nur die Möglichkeit $w = s$ bzw. $\bar{w} = \bar{s}$ übrig; die letztere Gleichung ist nach G 4 und Satz $1_1$ der ersteren äquivalent. Wenn also eventuell noch $w = s$ als Axiom zugefügt wird, ist keine nicht beweisbare Gleichheit mehr zufügbar. Hiermit ist der Hilfssatz bewiesen.

Gemäß (M 3) kommen für $mw$ nur *positive* Werte in Frage, an kürzeren Produkten also nur die Grundwerte $z$, $m$, $g$ und die Zweierprodukte $w$, $s$.

Die Forderung $mw = z$ würde nun aber auf

$$m = mz \text{ (Ia)} = mmw = mw \text{ (III a 1)} = z$$

führen, was nach der 1. Folgerung zu Satz 11 nicht sein kann.

Die Forderung $mw = g$ liefert (über $\overline{m}w = \overline{g}$) zunächst die Kommutativität $\overline{g}\overline{m} = \overline{m}\overline{g}$, nämlich:

$$\overline{g}\overline{m} = \overline{m}w\overline{m} = \overline{m}\overline{m}\overline{m}\overline{m} \text{ (D 1)} = \overline{m}\overline{m}w \text{ (D1)} = \overline{m}\overline{g};$$

hiermit ist $m = \overline{g}\overline{m}$ (III b 2) $= \overline{m}\overline{g} = g$ (III b 4), entgegen (M 4).

Daher bleiben höchstens solche zweistufigen idempotenten Modalitätenstrukturen, die sich durch die Forderungen $mw = m$, $= w$ bzw. $= s$ charakterisieren lassen.

*Axiomablengruppe VI.*

| | |
|---|---|
| a 1. $mw = m$, | a 2. $gs = g$, |
| b 1. $sm = m$, | b 2. $wg = g$, |
| c 1. $w\overline{m} = \overline{m}$, | c 2. $s\overline{g} = \overline{g}$, |
| d 1. $g\overline{w} = \overline{m}$, | d 2. $m\overline{s} = \overline{g}$ |

samt den durch Verneinung hervorgehenden Gleichheiten a 3—d 3 und a 4—d 4.

*Axiomablengruppe VII*

| | |
|---|---|
| a 1. $mw = w$, | a 2. $gs = s$, |
| b 1. $sm = w$, | b 2. $wg = s$, |
| c 1. $w\overline{m} = \overline{w}$, | c 2. $s\overline{g} = \overline{s}$, |
| d 1. $g\overline{w} = \overline{w}$, | d 2. $m\overline{s} = \overline{s}$, |

*Axiomablengruppe VIII.*

| | |
|---|---|
| a 1. $mw = s$, | a 2. $gs = w$, |
| b 1. $sm = s$, | b 2. $wg = w$, |
| c 1. $w\overline{m} = \overline{s}$, | c 2. $s\overline{g} = \overline{w}$, |
| d 1. $g\overline{w} = \overline{s}$, | d 2. $m\overline{s} = \overline{w}$ |

samt den durch Verneinung hervorgehenden Gleichheiten a 3—d 3 und a 4—d 4.

*Anmerkungen.* 1. Die Axiomablen sind hier unter Heranziehung der neuen Werte $w$, $s$ ausgedrückt. Bei Beschränkung auf die Grundwerte würde man die folgende Gestalt der Axiomablen haben:

| | VI. | VII. | VIII. | | VI. | VII. | VIII. |
|---|---|---|---|---|---|---|---|
| e 1. $m\overline{m}\overline{m} =$ | $m$ | $w$ | $s$ | e 2. $g\overline{g}\overline{g} =$ | $g$ | $s$ | $w$ |
| f 1. $mgm =$ | $m$ | $w$ | $s$ | f 2. $gmg =$ | $g$ | $s$ | $w$ |
| g 1. $\overline{g}m\overline{m} =$ | $m$ | $w$ | $s$ | g 2. $\overline{m}g\overline{g} =$ | $g$ | $s$ | $w$ |
| h 1. $\overline{g}\overline{g}m =$ | $m$ | $w$ | $s$ | h 2. $\overline{m}\overline{m}g =$ | $g$ | $s$ | $w$ |

samt den durch Verneinung hervorgehenden Gleichheiten e 3 — h 3 und e 4 — h 4.

2. Die Systeme der Gleichheiten III, VI sowie III, VII und III, VIII geben für jedes Produkt aus drei echten Grundwerten einen echten Grundwert oder ein Zweierprodukt aus Grundwerten an. In allen drei Fällen hat man zusammen mit I, II für jedes Produkt aus drei Grundwerten eine Reduktion auf einen kürzeren Wert; falls es sich um Modalitätenstrukturen handelt, sind sie höchstens zweistufig.

3. Das Axiom VI e 3 ist von der Gestalt des Ax. $\Sigma\,1$ mit $p = 1$.

(Bemerkungen zur *Interpretation* der Axiomablengruppen VI—VIII werden demnächst an anderer Stelle folgen.)

*Satz 21.* Im Rahmen der Axiome I, II allein (und der Regeln G sowie der Definition *D*) sind alle 16 Axiomablen der Gruppe VI, ebenso alle 16 Axiomablen der Gruppe VII und endlich alle 16 Axiomablen der Gruppe VIII einander äquivalent; es genügt also, *eine* Gleichheit aus einer dieser Gruppen dem Axiomensystem I—III als Axiom anzufügen.

Dies ist eine Folge des Quaternalitätssatzes und der Gleichheiten (T) aus Satz 20 (für die Axiomablen a—d) bzw. der Gleichheiten (Q) aus Satz 8 (für die Axiomablen e—h).

*Satz 22.* Bei Heranziehung der Definition D folgen:

1. alle Axiomablen VI aus jedem der Axiomablen IV oder V,
2. alle Axiomablen VII aus jedem der Axiomablen IV,
3. alle Axiomablen VIII aus jedem der Axiomablen V.

*Beweis*: Quaternalitätssatz 7. Dazu:

für 1. VI e 1 folgt aus IV.: $m\overline{m}\overline{m} = \overline{m}\overline{m}$ (IV a 1) $= m$ (IV a 2), (*)

VI e 1 folgt aus V.: $m\overline{m}\overline{m} = mg$ (V a 2) $= m$ (V b 3), (**)

für 2. $w = \overline{m}\overline{m}$ (D 1) $= m$ (IV a 2) $= m\overline{m}\overline{m}$ (nach (*)) $= mw$ (D 1),

für 3. $s = \overline{g}\overline{g}$ (D 2) $= m$ (V a 4) $= m\overline{m}\overline{m}$ (nach (**)) $= mw$ (D 1).

*Satz 23.* Die Axiomablen VI, VII, VIII sind von den Axiomablen I bis III im Rahmen der Schlußregeln G unabhängig.

Dies folgt unmittelbar auf Grund der Anmerkung 2 S. 85 und des Satzes $12_2$.

*Satz 24.* Jedes der 48 in VI, VII, VIII angeführten Axiomablen a bis d führt im Rahmen der Axiome I—III (und Regeln G) auf das Gleichheitsquadrupel

$$ww = w, \quad \overline{w}w = \overline{w}, \quad ss = s, \quad \overline{s}s = \overline{s}.$$

*Beweis.* $ww = \overline{m}\overline{m}\overline{m}\overline{m}$ (D 1) $= gnw\overline{m}$ (II und D 1) $= g\overline{w}\overline{m} = gmw$ (Gleichung (T) aus Satz 20) $\begin{cases} = gm \text{ (VI a 1)} = w \text{ (D 1)}; \\ = gw \text{ (VII a 1)} = w \text{ (U)}; \\ = gs \text{ (VIII a 1)} = w \text{ (VIII a 2)}. \end{cases}$

Dazu der Quaternalitätssatz.

*Satz 25.* Jede der Axiomenbasen I—III, VI bzw. I—III, VII bzw. I—III, VIII — genauer: jede Axiomenbasis, die aus den Axiomen I, je einem der Axiomablen II, III und weiter aus einem der Axiomablen VI, VII oder VIII besteht — spannt mit den Schlußregeln G eine zweistufige idempotente Modalitätenstruktur auf, die außer den sechs Grundwerten noch die vier Werte $w$, $\overline{w}$, $s$, $\overline{s}$ (und sonst keine) zu Basiswerten hat.

*Bezeichnung.* Die Gesamtheiten von Gleichheiten, die aus den genannten Axiomenbasen entfließen, seien „*die erste, zweite bzw. dritte Zehnmodalitätenstruktur*" genannt.

*Beweis.* Nach Satz 10 bleibt nur zu zeigen, daß die zehn Werte $z$, $\overline{z} = n$, $m$, $\overline{m}$, $g$, $\overline{g}$, $w$, $\overline{w}$, $s$, $\overline{s}$ nicht auf Grund der Axiome und Schlußregeln gleich werden können.

a) Wie im Beweise zu Satz 12 erhält man: Ist $\alpha = \beta$ auf Grund der ersten bzw. zweiten bzw. dritten Axiomenbasis und der Regeln G beweisbar, so ist in einer Halbgruppe mit den Elementen $m$ und $n$ das Wort $R\,\alpha$ dem

Wort $R\,\beta$ gleich ($R$ die $m$-$n$-Reduzierte S. 78), sofern man als definierende Relationen zugrundelegt: R 1. $nn = 1$, R 2. $mm = m$ und

im ersten Falle R $3_1$: $mnmnm = m$ (anschließend an VI e 1),
im zweiten Falle R $3_2$: $mnmnm = nmnm$ (anschließend an VII e 1 und D 1),
im dritten Falle R $3_3$: $mnmnm = mnmn$ (anschließend an VIII e 1 und D 2).

Es bleibt nun zu zeigen:

b) In den genannten Halbgruppen sind die zehn Reduzierten 1, $n$, $m$, $nm$, $nmn$, $mn$, $nmnm$, $mnm$, $mnmn$, $nmnmn$ nicht ineinander überführbar.

Beweis von b in der ersten Halbgruppe.

1. Ein Wort mit einer ungeraden Anzahl echter $n$-Abteilungen (S. 79) kann nur wieder in ein Wort mit einer ungeraden Anzahl echter $n$-Abteilungen übergehen.

2. Die Echtheit bzw. Unechtheit der einleitenden und ebenso der abschließenden $n$-Abteilungen ändert sich nicht.

3. Ein Wort, das $m$-Zeichen enthält, geht stets in ein Wort derselben Eigenschaft über.

1., 2. und 3. führen auf die Behauptung b.

Beweis von b in der zweiten Halbgruppe.

1′. Die Anzahl echter $n$-Abteilungen bleibt bei jeder Umformung eines Wortes erhalten.

2′. Die Echtheit bzw. Unechtheit der abschließenden $n$-Abteilungen ändert sich nicht.

3′. Ein Wort, das höchstens eine $m$-Abteilung enthält, ist nur nach R 1, R 2 umformbar und daher nur in ein Wort mit derselben Anzahl (0 bzw. 1) von $m$-Abteilungen überführbar.

1′. 2′. und 3′. führen auf die Behauptung b.

Beweis von b. in der dritten Halbgruppe entsprechend. Es ist lediglich in 2′. die einleitende statt der abschließenden $n$-Abteilung heranzuziehen.

*Satz* 26. Die Forderung VI e 3: $\overline{m}\overline{m}\overline{m} = \overline{m}$, die die Gestalt des Axioms $\Sigma$ 1 mit $p = 1$ hat (Anm. 3, S. 85), ist im Rahmen der Axiome I—III und der Regeln G nicht durch $\overline{m}\overline{m} = z$ oder durch $\overline{m}\overline{m}x = x$ ersetzbar.

*Beweis.* Mit $\overline{m}\overline{m}x = x$ ist $\overline{m}\overline{m} = z$, also nach (Q 2) des Satzes 8: $gm = z$ und weiter nach Satz 7: $mg = z$. Dies ist nach einem Schluß in Abs. 2 des Beweises zu Satz 13 für Modalitätenstrukturen ausgeschlossen. – Nun Satz 25.

*Satz* 27. — $27_1$. Fügt man den Axiomen der zweiten oder der dritten Zehnmodalitätenstruktur — d. h. den Axiomen I, je einem Axiomabel II, III und einem Axiomabel VII *oder* VIII — noch als Axiom IX die Gleichheit $w = s$ an, so erhält man eine zweistufige idempotente Modalitätenstruktur, die außer den sechs Grundwerten nur noch die beiden Grundwerte $w$, $\overline{w}$ zu Basiswerten hat. — $27_2$. Diese Modalitätenstruktur wird ebenso von der Axiomenbasis I—III, VII, VIII — genauer: von den Axiomen I und je einem Axiomabel II, III, VII, VIII — aufgespannt.

*Bezeichnung*: Die so eingeführte zweistufige Modalitätenstruktur heiße die „*Achtmodalitätenstruktur*“.

*Beweis.* $27_2$ ist eine unmittelbare Folge von $27_1$, da die Axiome VIIa 1, VIIIa 1 auf $w = s$ führen und umgekehrt aus jedem von ihnen mit $w = s$ auch das andere folgt.

Beweis von $27_1$. Nach Satz 10 bleibt nur zu zeigen, daß die acht Werte $z, \bar{z} = n, m, \bar{m}, g, \bar{g}, w\ (= s), \bar{w}\ (= \bar{s})$ nicht auf Grund der Axiome und Schlußregeln gleich werden können.

a) Wie im Beweise zu Satz 12 erhält man: Ist $\alpha = \beta$ auf Grund der in Satz $27_1$ aufgeführten Axiome und der Regeln G beweisbar, so ist in einer Halbgruppe mit den Elementen $m$, $n$ das Wort $R\,\alpha$ dem Wort $R\,\beta$ gleich ($R$ die $m$-$n$-Reduzierte S. 78), sofern man als definierende Relationen zugrunde legt:

R 1‴. $nn = 1$, R 2‴. $mm = m$, R 3‴. $mnmnm = nmnm = mnmn$.

Es bleibt zu zeigen:

b) In der genannten Halbgruppe sind die acht Reduzierten $1, n, m, nm, nmn, mn, nmnm\ (= mnmn), mnm$ nicht ineinander überführbar.

1. Die Anzahl echter $n$-Abteilungen bleibt bei jeder Umformung eines Wortes gerade bzw. ungerade.

Ein Wort, das höchstens eine $m$-Abteilung enthält, ist nur nach R 1‴., R 2‴. umformbar; daher ändert sich bei Umformung eines *solchen* Wortes

2. weder die Anzahl (0 bzw. 1) der abschließenden und der einleitenden echten $n$-Abteilungen

3. noch die Anzahl (0 bzw. 1) der $m$-Abteilungen.

Aus 2., 3. folgt, daß die ersten sechs genannten Reduzierten (die den Grundwerten entsprechen) voneinander und von den beiden letzten Reduzierten verschieden sind. Nach 1. sind die beiden letzten Reduzierten voneinander verschieden.

*Satz 28.* — $28_1$. Die vier in Satz 25 und Satz 27 eingeführten Modalitätenstrukturen sind die einzigen zweistufigen idempotenten Modalitätenstrukturen. — $28_2$. Keine zwei von ihnen sind isomorph. — $28_3$. Von den drei Zehnmodalitätenstrukturen ist keine in der anderen enthalten. Die erste Zehnmodalitätenstruktur ist nicht in der Achtmodalitätenstruktur enthalten. (Daß die Achtmodalitätenstruktur in keiner Zehnmodalitätenstruktur enthalten ist, ist trivial; ebenso ist nach $27_2$ evident, daß die zweite und dritte Zehnmodalitätenstruktur in der Achtmodalitätenstruktur enthalten sind.)

Beweis für $28_1$. Nach dem Hilfssatz dieses Paragraphen käme außer den eingeführten Modalitätenstrukturen zunächst nur noch eine solche in Betracht, die aus der *ersten* Zehnmodalitätenstruktur durch Zufügung von $w = s$ entsteht. Für diese würde aber gelten: $gm = w$ (D 1) $= gw$ (Gleichung U) $= gs = g$ (Ax. VIa 2). D. h. Ax. Vb 1 ist ableitbar; die Struktur ist einstufig.

Beweis für $28_2$.

1. Die erste Zehnmodalitätenstruktur stimmt nicht mit der zweiten überein:

Würden die beiden genannten Strukturen übereinstimmen, so würde aus VI a 1 und VII a 1 folgen $m = w$, d. h. nach D 1: $m = gm$, d. i. Ax. IV b 1. Die betreffende Struktur wäre einstufig, im Widerspruch zu Satz 25.

2. Die erste Zehnmodalitätenstruktur stimmt nicht mit der dritten überein:

Würden die beiden genannten Strukturen übereinstimmen, so würde aus VI a 1, VIII a 1 und D 2 folgen: $m = mg$, d. i. Ax. V b 3. Die betreffende Struktur wäre einstufig, entgegen Satz 25.

3. Die zweite Zehnmodalitätenstruktur stimmt nicht mit der dritten überein:

Würden diese Strukturen nämlich übereinstimmen, so würde nach Ax. VII a 1, VIII a 1 folgen: $w = s$, im Widerspruch zu der in Satz 25 angegebenen Wertebasis, die die Werte $w$ und $s$ als verschieden enthält.

4. Keine der Zehnmodalitätenstrukturen stimmt — wiederum wegen der Verschiedenheit von $w$ und $s$ — mit der Achtmodalitätenstruktur überein.

Nach Satz 13 sind daher keine zwei von den vier betrachteten Strukturen isomorph.

Beweis für $28_3$. Unter der Annahme, daß eine der Zehnmodalitätenstrukturen in einer anderen enthalten sei, gelangt man für die enthaltende Struktur wie in den Abs. 1—3 des Beweises für $28_2$ zum Widerspruch. — Die Annahme, daß die erste Zehnmodalitätenstruktur in der Achtmodalitätenstruktur enthalten sei, führt (da ja die zweite Zehnmodalitätenstruktur in ihr enthalten ist) zum selben Widerspruch.

## Literaturverzeichnis.

(1) MacColl, H.: Symbolic logic and its applications, London 1906. — (2) Lewis, C. I.: A survey of symbolic logic, Berkeley 1918, Kap. V. — (3) Łukasiewicz, Jan: C. r. Soc. Sci. et Lettres Warschau Cl. III, **23**, 51 (1930); vgl. auch den Bericht in Ruch Filosof. Lwow; **5**, Nr. 9, 169 (1920). — (4) Becker, O.: Jb. Philos. u. phänomen. Forsch. Halle, **11**, 497 (1930). — (5) Lewis, C. I. a. C. H. Langford: Symbolic logic, New York a. London 1932, Appendix II. — (6) Feys, R.: Rev. néoscolast. de philos. **40**, 517 (1937); **41**, 217 (1938). — (7) Churchman, C. W.: J. Symbol. Log. **3**, 77 (1938). — (8) Parry, W. T.: J. Symbol. Log. **4**, 137 (1939). — (9) Becker, O.: Bl. dtsch. Philos. **16**, 387 (1943).

*(Eingegangen am 24. Juni 1949.)*

Math. Annalen, Bd. 122, S. 90—108 (1950).

# Modulformen zweiten Grades und Dirichletreihen.

Von

HANS MAASS in Heidelberg.

Der vorliegende Aufsatz enthält erste Ansätze, um die von C. L. SIEGEL[1]) eingeführten Modulformen höheren Grades in die HECKEsche Theorie[2]) der DIRICHLETschen Reihen, die gewissen Funktionalgleichungen genügen, einzubeziehen. Da es mir zweckmäßig erscheint, zunächst einmal den einfachsten nicht-trivialen Fall gesondert darzustellen, beschränke ich mich hier auf die Betrachtung der Modulformen zweiten Grades. Es zeigt sich, daß jeder solchen Modulform mit Hilfe der MELLINschen Integraltransformation ein System von Dirichletreihen zugeordnet werden kann. Diese lassen sich in die ganze komplexe Ebene analytisch fortsetzen und genügen erwartungsgemäß gewissen Funktionalgleichungen. Die Art des Vorgehens wird durch zahlreiche Analogien zwischen den HILBERTschen und SIEGELschen Modulformen nahegelegt; d. h. man braucht nur die Methode, mit der E. HECKE[3]) aus den Thetareihen der reellen quadratischen Zahlkörper die zugehörigen Zetafunktionen mit Größencharakteren abgeleitet hat, auf die Modulformen zweiten Grades sinngemäß zu übertragen. An Stelle der Größencharaktere treten jetzt Funktionen, die sich aus gewissen automorphen Wellenfunktionen der hyperbolischen Ebene in ähnlicher Weise ableiten lassen wie die Größencharaktere aus der Exponentialfunktion. Die Frage, ob die einer Modulform zweiten Grades zugeordneten Reihen diese eindeutig bestimmen, bleibt offen. Hierzu bedarf es noch einiger Hilfsmittel aus der Eigenwerttheorie partieller Differentialgleichungen, die offenbar nicht ganz billig zu haben sind.

Wir bringen die wichtigsten Schritte, die zur Aufstellung der Zetafunktionen mit Größencharakteren führen, in Erinnerung, um die Analogie mit den folgenden Entwicklungen sinnfällig vor Augen zu haben.

Es sei $K$ ein reeller quadratischer Zahlkörper mit der Grundeinheit $\varepsilon > 1$ und $\mu \to \mu'$ der von der Identität verschiedene Automorphismus des Körpers $K$. Bekanntlich stellt die Thetareihe

$$(1) \qquad \vartheta(\tau, \tau') = \sum_{\mu} e^{ic(\mu^2\tau + \mu'^2\tau')},$$

in der $c$ eine gewisse von $K$ abhängige Konstante bezeichnet und $\mu$ gewisse ganze Zahlen in $K$ durchläuft, eine Modulform zu einer Untergruppe der HILBERTschen Modulgruppe des Körpers $K$ dar. An Stelle von $y = \mathfrak{Im}\,\tau$, $y' = \mathfrak{Im}\,\tau'$ führt HECKE die Variablen $u, v$ durch die Gleichungen

$$(2) \qquad y = u\, e^{2v\log\varepsilon}, \quad y' = u\, e^{-2v\log\varepsilon}$$

ein, um eine Invarianz gegenüber der Transformation $(y, y') \to (y\varepsilon^2, y'\varepsilon'^2)$ mit dem Prinzip der Fourierentwicklung in Verbindung bringen zu können.

---

[1]) SIEGEL, C. L.: Math. Ann. **116**, 617 (1939).
[2]) HECKE, E.: Math. Ann. **112**, 664 (1936).
[3]) HECKE, E.: 1. Mitt. Math. Z. **1**, 357 (1918); 2. Mitt. Math. Z. **6**, 11 (1920).

Alsdann liefert die Mellintransformation die in $v$ periodische Funktion

$$\begin{aligned}(3)\qquad \xi(s, v) &= \int_0^\infty [\vartheta(i y, i y') - a_0]\, u^{s-1}\, du \\ &= c^{-s}\,\Gamma(s) \sum_{\mu} (\mu^2 e^{2v\log\varepsilon} + \mu'^2 e^{-2v\log\varepsilon})^{-s}\end{aligned}$$

Hierin ist $a_0$ das konstante Glied der Thetareihe. $\xi(s, v)$ hat die Periode 1:

$$\xi(s, v+1) = \xi(s, v)$$

und stellt eine Epsteinsche Zetafunktion dar, wenn man vom Faktor $c^{-s}\,\Gamma(s)$ absieht. Die Fourierkoeffizienten $\xi_n(s)$ der Funktion $\xi(s, v)$ bestimmen diese, also auch $\vartheta(\tau, \tau')$ umkehrbar eindeutig; denn es ist

$$\vartheta(i y, i y') - a_0 = \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} \xi(s, v)\, u^{-s}\, ds\,.$$

Man erhält explizit

$$(4)\qquad \xi_n(s) = \int_0^1 \xi(s, v)\, e^{-2\pi i n v}\, dv = A^s\,\Gamma(s, n) \sum_{(\mu)} \frac{\lambda^n(\mu)}{|N\mu|^s}$$

mit

$$(5)\qquad \lambda(\mu) = \left|\frac{\mu}{\mu'}\right|^{\frac{\pi i}{\log\varepsilon}}$$

und einem gewissen elementaren Faktor $A^s\,\Gamma(s, n)$. Die Summation ist über gewisse Hauptideale $(\mu)$ zu erstrecken. Die Reihe

$$(6)\qquad \sum_{(\mu)} \frac{\lambda^n(\mu)}{|N\mu|^s}$$

stellt eine Zetafunktion von $K$ einfachster Art zum Größencharakter $\lambda^n$ dar.

Bei der Durchführung der entsprechenden Theorie für die Modulformen zweiten Grades ist ein Umstand von entscheidender Bedeutung, der in der Heckeschen Theorie[3]) kein Analogon besitzt: Um zu einer gegebenen Modulform zweiten Grades Dirichletreihen mit Hilfe der Mellintransformation konstruieren zu können, genügt es nicht, wenn man aus der Fourierentwicklung der Modulform das konstante Glied (entsprechend $a_0$ in (3)) herausnimmt, sondern es sind auch alle diejenigen Glieder auszusondern, die zu Exponentenmatrizen vom Rang 1 gehören. Die ausgeschlossenen Koeffizienten haben auf die Bildung der Dirichletreihen einen erkennbaren Einfluß insofern, als sie die Residuen in den endlich vielen möglichen Polen erster Ordnung eindeutig bestimmen. Um die Art dieser Abhängigkeit festzustellen, ist von den Ergebnissen der Heckeschen Theorie[2]) weitgehend Gebrauch zu machen. Das ist nicht sonderlich überraschend, wenn man bedenkt, daß jeder Modulform zweiten Grades eine solche ersten Grades durch einen einfachen Grenzprozeß eindeutig zugeordnet werden kann, und wenn man ferner beachtet, daß die Fourierkoeffizienten dieser Modulform ersten Grades mit den Entwicklungskoeffizienten der Form zweiten Grades übereinstimmen, die bei der Aufstellung der Dirichletreihen auszulassen sind. Allgemein kann jetzt schon gesagt werden, daß die Theorie für die Modulformen $(n - 1)$-ten Grades vollständig durchgebildet sein muß,

wenn man zu Modulformen $n$-ten Grades übergehen will. D. h. man wird das Prinzip der vollständigen Induktion nach $n$ in wesentlichen Punkten anwenden müssen.

Im Raum der positiven symmetrischen Matrizen zweiten Grades erweist sich die Parameterdarstellung[4])

$$Y = \begin{pmatrix} u(x^2+y^2)\,y^{-1} & u\,x\,y^{-1} \\ u\,x\,y^{-1} & u\,y^{-1} \end{pmatrix} \qquad (u>0,\ y>0) \tag{7}$$

als geeignet, um zu einfachen analytischen Ausdrücken zu gelangen, mit denen die im Verlauf der Untersuchung notwendigen Operationen durchgeführt werden können. Die metrische Fundamentalform

$$ds^2 = \tfrac{1}{2}\,Sp\,(Y^{-1}d\,Y)^2 \tag{8}$$

die gegenüber den Transformationen

$$Y \to U\,Y\,U' \qquad (U \text{ reell, } |U| \neq 0)$$

bekanntlich invariant ist, nimmt hier die Gestalt

$$ds^2 = \frac{du^2}{u^2} + \frac{dx^2+dy^2}{y^2} \tag{9}$$

an. Mithin kann die Fläche $|Y| \equiv u^2 = 1$ als ein Modell der hyperbolischen Ebene angesehen werden. $U$ bezeichne eine reelle Matrix zweiten Grades mit der Determinante $+1$; ferner sei $\tau = x + i\,y$. Dann sind $Y \to U\,Y\,U'$ und $\tau \to U(\tau)$ verschiedene Darstellungen ein und derselben hyperbolischen Bewegung.

## § 1. Aufstellung der Dirichletreihen.

Wir bezeichnen mit großen lateinischen Buchstaben zweireihige quadratische Matrizen; insbesondere sei $E$ die Einheitsmatrix und $O$ die Nullmatrix. Genügen die ganzzahligen Matrizen $A, B, C, D$ den Beziehungen

$$A\,B' = B\,A', \quad C\,D' = D\,C', \quad A\,D' - B\,C' = E,$$

dann stellt

$$\sigma = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

eine Modulsubstitution zweiten Grades dar. Die Gesamtheit dieser Substitutionen bildet bezüglich der Matrizenmultiplikation eine Gruppe $\mathsf{M}_2$, die SIEGELsche Modulgruppe zweiten Grades. Der Raum $\mathfrak{H}$ der symmetrischen komplexen Matrizen zweiten Grades mit positivem Imaginärteil:

$$Z = \begin{pmatrix} z_0 & z_1 \\ z_1 & z_2 \end{pmatrix} = X + i\,Y, \quad X = \begin{pmatrix} x_0 & x_1 \\ x_1 & x_2 \end{pmatrix}, \quad Y = \begin{pmatrix} y_0 & y_1 \\ y_1 & y_2 \end{pmatrix} > 0 \tag{10}$$

wird durch die Substitutionen $\sigma \subset \mathsf{M}_2$ vermöge

$$\sigma(Z) = (A\,Z + B)\,(C\,Z + D)^{-1} \tag{11}$$

in sich übergeführt. Sei allgemein

$$R\,[Q] = Q'\,R\,Q = [Q']\,R \text{ und } \operatorname{abs}(R) = \text{Betrag der Determinante } |R|,$$

Offenbar haben die Matrizen $(T\,[U])\,Y$ und $T\,([U]\,Y)$ dieselbe Spur; denn sie gehen durch zyklische Vertauschung der Faktoren auseinander hervor.

[4]) s. auch C. L. SIEGEL: Amer. J. Math. **65**, 1 (1943), insbesondere Lemma 6.

$Sp\,(T\,[U]\,Y)$ ist also unabhängig davon, ob $[U]$ als Rechtsoperator von $T$ oder Linksoperator von $Y$ aufgefaßt wird.

Ein Fundamentalbereich $\mathfrak{F}_2$ von $\mathsf{M}_2$ in $\mathfrak{H}$ wird nach SIEGEL[1]) durch folgende Ungleichungen beschrieben:

1. $\mathrm{abs}\,(CZ + D) \geqq 1$ für alle zweiten „Matrizenzeilen" $C, D$ von Modulsubstitutionen zweiten Grades.

2. $0 \leqq 2\,y_1 \leqq y_2 \leqq y_0$ (MINKOWSKIsche Reduktionsbedingungen).

3. $-\frac{1}{2} \leqq x_\nu \leqq \frac{1}{2}$ für $\nu = 0, 1, 2$.

Es sei $k$ eine natürliche Zahl und $v\,(\sigma)$ ein abelscher Charakter von $\mathsf{M}_2$. Unter einer Modulform (zweiten Grades) der Dimension $-k$ zum Multiplikatorsystem $v\,(\sigma)$ verstehen wir eine Funktion $g\,(Z)$, die in $\mathfrak{H}$ eine reguläre Funktion von $z_0, z_1, z_2$ darstellt, in $\mathfrak{F}_2$ beschränkt ist und der Transformationsformel

(12) $$g\,(\sigma\,(Z)) = v\,(\sigma)\,|\,CZ + D\,|^k\, g\,(Z) \qquad \text{für } \sigma = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \subset \mathsf{M}_2$$

genügt. Wenn $g\,(Z)$ nicht identisch verschwindet, dann ist notwendig

$$v\begin{pmatrix} -E & O \\ O & -E \end{pmatrix} = 1\,.$$

Wir beschränken uns auf solche Charaktere und fordern überdies noch

$$v\begin{pmatrix} E & S \\ O & E \end{pmatrix} = v\begin{pmatrix} U' & O \\ O & U^{-1} \end{pmatrix} = 1$$

für ganzzahlige $S, U$ mit $S = S'$ und $|\,U\,| = 1$. Aus den Relationen

$$\begin{pmatrix} O & E \\ -E & O \end{pmatrix}^2 = \begin{pmatrix} -E & O \\ O & -E \end{pmatrix} \quad \text{und} \quad \left\{\begin{pmatrix} O & E \\ -E & O \end{pmatrix}\begin{pmatrix} E & U \\ O & E \end{pmatrix}\right\}^3 = \begin{pmatrix} U & O \\ O & U \end{pmatrix} \quad \text{für } U^2 = E$$

ist

$$v\begin{pmatrix} O & E \\ -E & O \end{pmatrix}^2 = 1 \quad \text{und} \quad v\begin{pmatrix} O & E \\ -E & O \end{pmatrix} = v\begin{pmatrix} U' & O \\ O & U^{-1} \end{pmatrix}$$

für ganzzahlige $U$ mit $U = U' = U^{-1}$ zu entnehmen. Setzen wir speziell $U = E$ und $U = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, so ergibt sich

$$v\begin{pmatrix} O & E \\ -E & O \end{pmatrix} = v\begin{pmatrix} U_0' & O \\ O & U_0^{-1} \end{pmatrix} = 1 \qquad \text{mit } U_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Da die Modulsubstitutionen

$$\begin{pmatrix} E & S \\ O & E \end{pmatrix}, \ \begin{pmatrix} U' & O \\ O & U^{-1} \end{pmatrix}, \ \begin{pmatrix} O & E \\ -E & O \end{pmatrix} \qquad (S' = S\,,\,|\,U\,| = \pm 1)$$

nach E. WITT[5]) die Gruppe $\mathsf{M}_2$ erzeugen, so ist offenbar $v\,(\sigma) = 1$ für alle $\sigma \subset \mathsf{M}_2$. Es erweist sich also als unnötig, Multiplikatorsysteme einzuführen.

Jeder Modulform $g\,(Z)$ läßt sich eine Modulform ersten Grades $g^*\,(z_0)$ eindeutig zuordnen. Man erklärt sie als Grenzfunktion:

(13) $$g^*\,(z_0) = \lim_{\lambda \to \infty} g\begin{pmatrix} z_0 & 0 \\ 0 & i\lambda \end{pmatrix}$$

und beweist mit Hilfe von (12) die Transformationsformel

(14) $$g^*\,(U\,(z_0)) = (c\,z_0 + d)^k\, g^*\,(z_0)$$

[5]) WITT, E.: Hamburger Abh. 14, 323 (1941).

für beliebige Modulsubstitutionen ersten Grades $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Für $k \equiv 1$ (2) verschwindet $g^*(z_0)$ notwendig identisch.

Aus der Invarianz von $g(Z)$ gegenüber den Modulsubstitutionen $Z \to Z + S$ und der Beschränktheit in $\mathfrak{F}_2$ folgt die Möglichkeit einer FOURIERschen Reihenentwicklung

$$g(Z) = \sum_T a(T)\, e^{2\pi i Sp(TZ)}, \tag{15}$$

in der über alle halbganzen nicht-negativen symmetrischen $T$ summiert wird. Für ganzzahlige $U$ mit $|U| = 1$ gilt

$$a(T[U]) = a(T). \tag{16}$$

$T$ und $T^*$ sollen eigentlich assoziiert heißen, wenn es eine ganzzahlige Matrix $U$ mit der Determinante 1 gibt, so daß $T^* = T[U]$ wird. $\{T\}$ bezeichne einen Repräsentanten in der Schar der zu $T$ eigentlich assoziierten Matrizen.

Hat $T$ den Rang 1, so gibt es einen Repräsentanten $\{T\} = \begin{pmatrix} 0 & 0 \\ 0 & t \end{pmatrix}$ $(t > 0)$. Wir setzen $a_t = a(T)$ für $T = \begin{pmatrix} 0 & 0 \\ 0 & t \end{pmatrix}$, $t \geqq 0$. Die Koeffizienten $a_t$ bestimmen $g^*(z_0)$ eindeutig:

$$g^*(z_0) = \sum_{t=0}^{\infty} a_t\, e^{2\pi i t z_0}. \tag{17}$$

Ist der Rang von $T$ gleich 2, so gibt es nur endlich viele ganzzahlige $U$ mit $|U| = 1$ und $T[U] = T$; ihre Anzahl sei $\varepsilon(T)$. Offenbar ist $\varepsilon(T) = \varepsilon(\{T\})$.

Wir führen nun die Parameterdarstellung

$$Y = u\, Y_1, \quad Y_1 = \begin{pmatrix} (x^2 + y^2)\, y^{-1} & x\, y^{-1} \\ x\, y^{-1} & y^{-1} \end{pmatrix} \tag{18}$$

ein und setzen $\tau = x + i y$. Der Raum der positiven $Y$ wird durch die Ungleichungen $u > 0$, $y > 0$ beschrieben. Besteht zwischen $Y$ und $u, \tau$ die angegebene umkehrbar eindeutige Beziehung, was wir kurz durch $Y \leftrightarrow (u, \tau)$ zum Ausdruck bringen, so folgt auch

$$[U]\, Y \leftrightarrow (u, U(\tau)) \quad \text{und} \quad Y^{-1} \leftrightarrow \left(\frac{1}{u}, -\frac{1}{\tau}\right), \tag{19}$$

was mühelos bestätigt werden kann. Dabei ist $U$ eine beliebige ganzzahlige Matrix mit der Determinante 1. Für die Gruppe der Transformationen

$$Y \to [U]\, Y \qquad (U \text{ ganzzahlig}, |U| = 1)$$

stellt der Bereich

$$\mathfrak{F}: \begin{cases} 0 \leqq |2\, y_1| \leqq y_2 \leqq y_0, & 1 \leqq u \quad \text{oder} \\ 0 \leqq |2\, y_1| \leqq y_0 \leqq y_2, & 0 < u \leqq 1 \end{cases} \tag{20}$$

einen Fundamentalbereich im Raum der positiven $Y$ dar. Als Fundamentalbereich der Modulgruppe ersten Grades wählen wir in der oberen $\tau$-Halbebene die Modulfigur

$$\mathfrak{F}_1: |2\, x| \leqq 1, \quad x^2 + y^2 \geqq 1, \quad y > 0. \tag{21}$$

Man erkennt sofort die Gleichwertigkeit der Aussagen:

$$Y \subset \mathfrak{F} \rightleftarrows \tau \subset \mathfrak{F}_1\,,\ 1 \leqq u \quad \text{oder} \quad -\frac{1}{\tau} \subset \mathfrak{F}_1\,,\ 0 < u \leqq 1\,. \tag{22}$$

Die Abbildung $Y \to Y^{-1}$ führt $\mathfrak{F}$ in sich über.

Wir setzen in der Fourierreihe (15) $X = O$ und nehmen eine Aufspaltung in Teilreihen vor:

$$f(Y) = g(i\,Y) = a_0 + f_1(Y) + f_2(Y)\,, \tag{23}$$

$$f_\nu(Y) = \sum_{\text{Rang } T = \nu} a(T)\, e^{-2\pi Sp(TY)} \qquad (\nu = 1, 2)\,. \tag{24}$$

Hat $T$ den Rang 1, so gibt es eine natürliche Zahl $t$ und eine ganzzahlige Matrix $U$ mit $|U| = 1$, so daß $T = \begin{pmatrix} 0 & 0 \\ 0 & t \end{pmatrix}[U]$ wird. Eine einfache Rechnung ergibt dann

$$Sp(T\,Y) = t\,u\,\frac{|c\tau + d|^2}{y}\,,$$

wenn $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ angenommen wird. Damit erhält man schließlich

$$f_1(Y) = \frac{1}{2}\sum_{t=1}^{\infty} a_t \sum_{(c,d)=1} e^{-2\pi t u \frac{|c\tau + d|^2}{y}}\,, \tag{25}$$

$$f_2(Y) = \sum_{\{T\}>0} \frac{a(T)}{\varepsilon(T)} \sum_{|U|=1} e^{-2\pi Sp(T[U]Y)}\,. \tag{26}$$

Summiert wird hier über alle Paare von teilerfremden ganzen Zahlen $c, d$ bzw. über alle ganzzahligen $U$ mit $|U| = 1$.

In analoger Weise behandeln wir die Fourierreihe (17):

$$f^*(y_0) = g^*(i\,y_0) = a_0 + f_1^*(y_0)\,, \tag{27}$$

$$f_1^*(y_0) = \sum_{t=1}^{\infty} a_t\, e^{-2\pi t y_0}\,. \tag{28}$$

Auf Grund der angegebenen Reihenentwicklung kann

$$f_1(Y) = \frac{1}{2}\sum_{(c,d)=1} f_1^*\left(u\,\frac{|c\tau + d|^2}{y}\right) \tag{29}$$

festgestellt werden. Ferner ist

$$f_\nu(Y) = f_\nu(u\,Y_1) = f_\nu(u\,Y_1^{-1}) \qquad (\nu = 1, 2)\,. \tag{30}$$

Das folgt sofort aus (24). Ersetzt man nämlich $T$ durch $T\left[\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\right]$, so ändert sich $a(T)$ nicht, während

$$Sp(T\,Y) \quad \text{in} \quad Sp\left(T\left[\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\right] Y\right) = u\,Sp\left(T\,Y_1^{-1}\right)$$

übergeht. Schließlich vermerken wir noch die sich aus den Transformationsformeln

$$g(-Z^{-1}) = |Z|^k g(Z) \quad \text{und} \quad g^*(-z_0^{-1}) = z_0^k g^*(z_0)$$

ergebenden Beziehungen

$$f_2(Y^{-1}) = \gamma\,|Y|^k f_2(Y) + \{\gamma\,|Y|^k f_1(Y) - f_1(Y^{-1})\} + \{\gamma\,|Y|^k a_0 - a_0\}\,. \tag{31}$$

$$f_1^*(y_0^{-1}) = \gamma^* y_0^k f_1^*(y_0) + \{\gamma^* y_0^k a_0 - a_0\}. \tag{32}$$

Zur Abkürzung ist hier $\gamma = (-1)^k$ und $\gamma^* = i^k$ gesetzt worden.

Bevor wir den Übergang zu den DIRICHLETschen Reihen vollziehen, wollen wir noch zur Sicherung ihrer Konvergenz die Fourierkoeffizienten $a(T)$ von $g(Z)$ geeignet abschätzen. Im folgenden seien $c_1, c_2, c_3, c_4$ feste nur von $g(Z)$ abhängige positive Konstanten. Wir erklären die Diskriminante von $T$ durch

$$D(T) = \begin{cases} |T| & \text{für} \quad T > 0, \\ t & \text{für} \quad T = \begin{pmatrix} 0 & 0 \\ 0 & t \end{pmatrix}[U] \quad (U \text{ ganzzahlig}, |U| = 1) \end{cases} \tag{33}$$

und beweisen, daß bei geeigneter Wahl von $c_1$

$$|a(T)| < c_1 (D(T))^k \quad \text{für} \quad T \neq O \tag{34}$$

gilt. Hat $T$ den Rang 1, so ist $a(T)$ Fourierkoeffizient einer Modulform ersten Grades. Koeffizienten dieser Art lassen sich bekanntlich in der gewünschten Weise abschätzen, so daß wir nur noch den Fall $T > 0$ zu betrachten brauchen. Ferner dürfen wir voraussetzen, daß $T = \begin{pmatrix} t_0 & t_1 \\ t_1 & t_2 \end{pmatrix}$ die MINKOWSKIschen Reduktionsbedingungen

$$0 \leq |2t_1| \leq t_2 \leq t_0$$

erfüllt; denn weder $a(T)$ noch $|T|$ ändert sich, wenn man $T$ durch eine zu $T$ eigentlich assoziierte Matrix ersetzt. Es ist dann $t_0 t_2 \leq \frac{4}{3}|T|$, $1 \leq t_0$, $1 \leq t_2$, also

$$t_0 \leq \tfrac{4}{3}|T|, \quad t_2 \leq \tfrac{4}{3}|T|. \tag{35}$$

Zunächst schätzen wir $g(Z)$ ab. Dabei verwenden wir eine von H. BRAUN[6]) angegebene Schlußweise. Zu vorgegebenem $Z$ bestimmen wir $\sigma = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \subset \mathsf{M}_2$ so, daß $Z_1 = \sigma(Z)$ in den SIEGELschen Fundamentalbereich $\mathfrak{F}_2$ fällt. Es ist dann $|g(Z_1)| \leq c_2$. Wir wählen $\varrho = \begin{pmatrix} S & -E \\ E & O \end{pmatrix} \subset \mathsf{M}_2$, behalten uns eine nähere Bestimmung von $S$ vor und setzen $Z_0 = \varrho^{-1}(Z) = (-Z+S)^{-1}$. Damit wird $Z_1 = \sigma\varrho(Z_0)$ und $\sigma\varrho = \begin{pmatrix} * & * \\ CS+D & -C \end{pmatrix}$, also nach (12)

$$g(Z_0) = |-Z+S|^k g(Z),\ g(Z_1) = |(CS+D)Z_0 - C|^k g(Z_0).$$

Die Koeffizienten der Matrix $S = \begin{pmatrix} s_0 & s_1 \\ s_1 & s_2 \end{pmatrix}$ lassen sich im Bereich $|s_\nu - x_\nu| < 2$ $(\nu = 0, 1, 2)$ ganzzahlig so bestimmen, daß $|CS+D| \neq 0$, also dem Betrag nach mindestens gleich 1 wird; denn $|CS+D|$ ist ein nicht identisch verschwindendes Polynom in $s_0, s_1, s_2$ vom Grad 2. Bezeichnen wir den Imaginärteil von $Z_0$ mit $Y_0$, so ergibt sich

$$\begin{aligned} |g(Z)| &= \operatorname{abs}(-Z+S)^{-k} \operatorname{abs}((CS+D)Z_0 - C)^{-k} |g(Z_1)| \\ &\leq \operatorname{abs}(-Z+S)^{-k} \operatorname{abs}(Z_0 - (CS+D)^{-1}C)^{-k} c_2 \\ &\leq \operatorname{abs}(-Z+S)^{-k} |Y_0|^{-k} c_2 = \operatorname{abs}(-Z+S)^k |Y|^{-k} c_2. \end{aligned}$$

---

[6]) BRAUN, H.: Math. Z. 44, 387 (1939).

Da die Koeffizienten der Matrix $-X + S$ beschränkt sind, kann offenbar

$$|g(Z)| \leqq c_3 |Y|^{-k} (1 + Sp\, Y)^{2k} \tag{36}$$

geschlossen werden. Setzt man hierin $Y = T^{-1}$, so folgt unter Berücksichtigung von (35)

$$|g(X + i\,T^{-1})| \leqq c_4 |T|^k \,.$$

Aus der FOURIERschen Koeffizientenformel

$$a(T) = \int\!\!\int\!\!\int_0^1 g(Z)\, e^{-2\pi i Sp(TZ)}\, d x_0\, d x_1\, d x_2$$

ergibt sich somit

$$|a(T)| \leqq c_4 |T|^k e^{4\pi} = c_1 |T|^k, \text{ q. e. d.}$$

Wir wenden nun die Mellintransformation auf $f_\nu(Y)$ an, bilden also

$$\xi_\nu(s, \tau; g) = \int_0^\infty f_\nu(u\, Y_1)\, u^{2s-1}\, du \qquad (\nu = 1, 2)\,. \tag{37}$$

Trägt man die Entwicklung (24) hier ein, so ergibt gliedweise Integration

$$\xi_\nu(s, \tau; g) = (2\pi)^{-2s}\, \Gamma(2s)\, \varphi_\nu(s, \tau; g) \tag{38}$$

mit

$$\varphi_\nu(s, \tau; g) = \sum_{\text{Rang}\, T = \nu} a(T)\, (Sp(T\, Y_1))^{-2s} = \sum_{\text{Rang}\, T = \nu} a(T) \left(\frac{y}{(x^2 + y^2) t_0 + 2 x t_1 + t_2}\right)^{2s} \tag{39}$$

Dabei ist $T = \begin{pmatrix} t_0 & t_1 \\ t_1 & t_2 \end{pmatrix}$ angenommen worden. Diese Reihen sind in gewissem Sinne Analoga zu der EPSTEINschen Zetafunktion, die in (3) vorkommt. Das Hauptaugenmerk ist auf $\varphi_2$ zu richten. $\varphi_1$ kann von unserm Standpunkt aus als elementar angesehen werden. Zufolge

$$Sp(T\, Y_1) = t\, \frac{|c\tau + d|^2}{y} \quad \text{für} \quad T = \begin{pmatrix} 0 & 0 \\ 0 & t \end{pmatrix} \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right]$$

wird nämlich

$$\varphi_1(s, \tau; g) = G(\tau, 2s)\, \varphi^*(2s, g^*)\,, \tag{40}$$

wenn

$$G(\tau, s) = \frac{1}{2} \sum_{(c,d) = 1} \frac{y^s}{|c\tau + d|^{2s}} \quad \text{und} \quad \varphi^*(s, g^*) = \sum_{t=1}^{\infty} \frac{a_t}{t^s} \tag{41}$$

gesetzt wird. $\varphi^*$ stellt die der Modulform $g^*$ im Sinne der HECKEschen Theorie zugeordnete Dirichletreihe dar.

Ersetzt man in (39) $T$ durch $T[U]$ ($U$ ganzzahlig, $|U| = 1$), so ändert sich $\varphi_\nu$ offenbar nicht; gleichwertig damit ist aber auch die Transformation $Y_1 \to [U]\, Y_1$, woraus

$$\varphi_\nu(s, U(\tau); g) = \varphi_\nu(s, \tau; g) \tag{42}$$

für ganzzahlige $U$ mit $|U| = 1$ erhellt. Wir zeigen noch, daß $\varphi_2(s, \tau; g)$ für hinreichend große Werte von $\Re s$ gleichmäßig in $\tau$ konvergiert, wenn $\tau$ in $\mathfrak{F}_1$ variiert. Unter der Voraussetzung $T = \begin{pmatrix} t_0 & t_1 \\ t_1 & t_2 \end{pmatrix} > 0$, $\tau \subset \mathfrak{F}_1$ gilt die Abschätzung

$$Sp(T\,Y_1) = \frac{x^2+y^2}{y}\,t_0 + 2\,\frac{x}{y}\,t_1 + \frac{1}{y}\,t_2 \geqq \frac{x^2+y^2}{y}\,t_0 - \frac{\sqrt{x^2+y^2}}{y}\sqrt{t_0 t_2} + \frac{1}{y}\,t_2$$
$$\geqq \frac{1}{2}\left(\frac{x^2+y^2}{y}\,t_0 + \frac{1}{y}\,t_2\right) \geqq \frac{1}{2}\left(y\,t_0 + \frac{t_2}{y}\right) \geqq \sqrt{t_0 t_2}\,.$$

Mit Hilfe von (34) ergibt sich somit in der Tat

$$|\varphi_2(s,\tau;g)| \leqq c_1 \sum_{\substack{0<t_0,t_2\\ |t_1|<\sqrt{t_0t_2}}} |T|^k\,(t_0 t_2)^{-s} \leqq 3\,c_1 \sum_{t_0,t_2=1}^{\infty} (t_0 t_2)^{k+\frac{1}{2}-s}$$
$$= 3\,c_1\,\zeta(s-k-\tfrac{1}{2})^2;\quad \text{dabei ist}\quad \zeta(s) = \sum_{n=1}^{\infty} n^{-s}.$$

Die Reihe $\varphi_2(s,\tau;g)$ ist bei festem $s$ nicht nur in $\mathfrak{F}_1$, sondern wegen (42) auch in der Halbebene $y>0$ beschränkt. Ferner gilt $\lim\limits_{y\to\infty} \varphi_2(s,\tau;g)=0$; denn jedes Glied der Reihe konvergiert für $y\to\infty$ gegen 0.

Es liegt nun nahe, die Funktion $\varphi_2$ oder $\xi_2$ einer Fourieranalyse betreffs der Funktionen $e(\tau)$ zu unterziehen, die folgenden Bedingungen genügen:

1. $e(\tau)$ ist in der Halbebene $y>0$ zweimal stetig differenzierbar und beschränkt.

2. $e(\tau)$ genügt der Wellengleichung

(43) $$\left[y^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)+\tfrac{1}{4}+r^2\right]e(\tau)=0.$$

3. Für alle Substitutionen $U$ der Modulgruppe ersten Grades ist

(44) $$e(U(\tau)) = e(\tau).$$

Automorphe Eigenfunktionen dieser Art werden künftig der Normierung

(45) $$\frac{3}{\pi}\iint_{\mathfrak{F}_1} e(\tau)\,\overline{e(\tau)}\,\frac{dx\,dy}{y^2} = 1.$$

unterworfen. Ist $e(\tau)$ konstant, also $r=\pm\frac{i}{2}$, so ist $|e(\tau)|=1$; denn der hyperbolische Inhalt von $\mathfrak{F}_1$ ist gerade $\frac{\pi}{3}$. Es soll dann $e(\tau)=1$ angenommen werden. Wir bestimmen zunächst einige einfache Eigenschaften der $e(\tau)$.

Aus den aufgeführten Bedingungen folgt, daß $e(\tau)$ in eine Fourierreihe der Art

(46) $$e(\tau) = u(y) + \sum_{n\neq 0} b_n\,y^{\frac{1}{2}}\,K_{ir}(2\pi|n|\,y)\,e^{2\pi i n x}$$

entwickelt werden kann. Dabei ist

(47) $$u(y) = \begin{cases} b_0\,y^{\frac{1}{2}+ir} + c_0\,y^{\frac{1}{2}-ir} & \text{für}\quad r\neq 0,\\ b_0\,y^{\frac{1}{2}}\log y + c_0\,y^{\frac{1}{2}} & \text{für}\quad r=0.\end{cases}$$

Wir wenden den GREENschen Satz

(48) $$\iint_{\mathfrak{B}} (V(\tau)\,\Delta U(\tau) - U(\tau)\,\Delta V(\tau))\,dx\,dy = \int_{\mathfrak{R}} \left(V(\tau)\frac{\partial U(\tau)}{\partial n} - U(\tau)\frac{\partial V(\tau)}{\partial n}\right)ds,$$

wobei $\mathfrak{R}$ den Rand eines Bereiches $\mathfrak{B}$, $n$ die nach außen weisende Normale und $s$ die Bogenlänge auf $\mathfrak{R}$ bedeutet, auf $U(\tau) = e(\tau)$ und $V(\tau) = \overline{e(\tau)}$ an. $\mathfrak{B} = \mathfrak{B}_q$ sei speziell der Bereich $0 < y \leqq q$, $|2x| \leqq 1$, $x^2 + y^2 \geqq 1$. Für $q \to \infty$ geht $\mathfrak{B}_q$ offenbar in $\mathfrak{F}_1$ über. Zufolge (43) wird

$$(-r^2 + \bar{r}^2) \iint_{\mathfrak{B}_q} e(\tau)\,\overline{e(\tau)}\,\frac{dx\,dy}{y^2} = -2i\,\mathfrak{Im} \int_0^1 e(\tau)\,\frac{\partial \overline{e(\tau)}}{\partial y}\,dx \Big|_{y=q};$$

denn die Beiträge des Randintegrals zu äquivalenten Randstücken heben sich auf. Eine einfache Rechnung zeigt, daß

$$\int_0^1 e(\tau)\,\frac{\partial \overline{e(\tau)}}{\partial y}\,dx = u(y)\,\overline{u'(y)} + \sum_{n \neq 0} a_n \bar{a}_n\, y^{\frac{1}{2}} K_{ir}(2\pi|n|\,y)\,\frac{d}{dy}\left(y^{\frac{1}{2}} K_{i\bar{r}}(2\pi|n|\,y)\right)$$

ist. Aus bekannten Eigenschaften der Besselschen Funktion $K_\nu(z)$ geht hervor, daß die unendliche Reihe $\sum\limits_{n \neq 0}$ für $y \to \infty$ gegen 0 konvergiert. Mithin liefert der Grenzübergang $q \to \infty$:

$$\frac{\pi}{3}(-r^2 + \bar{r}^2) = -2i \lim_{y \to \infty} \mathfrak{Im}\, u(y)\,\overline{u'(y)}.$$

Setzt man $r = \alpha + i\beta$ ($\alpha$, $\beta$ reell), so ergibt eine elementare Umformung

$$\begin{aligned} \frac{2\pi}{3}\alpha\beta &= \lim_{y \to \infty} \mathfrak{Im}\, u(y)\,\overline{u'(y)} \\ &= \lim_{y \to \infty} \begin{cases} (c_0 \bar{c}_0 y^{2\beta} - b_0 \bar{b}_0 y^{-2\beta})\,\alpha + (-i b_0 \bar{c}_0 y^{2i\alpha} + i c_0 \bar{b}_0 y^{-2i\alpha})\,\beta & \text{für } r \neq 0, \\ \mathfrak{Im}\, c_0 \bar{b}_0 & \text{für } r = 0. \end{cases} \end{aligned}$$

Dieser Limes kann im Falle $\alpha\beta \neq 0$ nur dann existieren, wenn entweder $b_0$ oder $c_0$ gleich 0 ist. Dann ist der Limes selber gleich 0, was aber mit $\alpha\beta \neq 0$ im Widerspruch steht. Allgemein gilt also $\alpha\beta = 0$; d. h. $r^2$ ist reell. Nach der Fourierschen Koeffizientenformel ist ferner

$$\int_0^1 e(\tau)\, e^{-2\pi i n x}\, dx = b_n\, y^{\frac{1}{2}} K_{ir}(2\pi|n|\,y) \quad \text{für} \quad n \neq 0.$$

Dieser Ausdruck ist beschränkt. Sei $\beta \neq \pm\frac{1}{2}$, also $\frac{1}{4} + r^2 \neq 0$; dann verschwinden nicht alle $b_n$; mithin ist auch

$$y^{\frac{1}{2}} K_{ir}(y) \quad \text{für } y \to 0 \text{ beschränkt.}$$

Berücksichtigt man

$$\lim_{y \to 0+0} y^\nu K_{\pm\nu}(y) = 2^{\nu-1}\,\Gamma(\nu) \quad \text{für} \quad \mathfrak{Re}\,\nu > 0,$$

so ergibt sich offenbar $|\beta| < \frac{1}{2}$, mithin $\frac{1}{4} + r^2 > 0$. Der Fall $\beta = \pm\frac{1}{2}$ führt auf eine Potentialfunktion, die in jedem Punkt einer geschlossenen Fläche regulär ist. Eine solche Funktion ist notwendig konstant. Es ist also allgemein

$$\beta = 0 \quad \text{oder} \quad \alpha = 0,\ |\beta| \leqq \tfrac{1}{2}. \tag{49}$$

Da auch

$$\int_0^1 e(\tau)\,dx = u(y)$$

beschränkt ist, so folgt nunmehr (für $y \to \infty$) $b_0 = c_0 = 0$, d. h. $u(y) = 0$, sofern $e(\tau)$ nicht konstant, also $|\beta| < \frac{1}{2}$ ist. Setzen wir generell

$$\delta(e) = \begin{cases} 1 \text{ für } e(\tau) = \text{konstant}, \\ 0 \text{ für } e(\tau) \neq \text{konstant}, \end{cases} \tag{50}$$

dann ist also

$$u(y) = \delta(e). \tag{51}$$

Wir wenden (48) auf $U(\tau) = e(\tau)$ und $V(\tau) = 1$ an und erhalten

$$-(\tfrac{1}{4} + r^2) \iint_{\mathfrak{B}_q} e(\tau) \frac{dx\,dy}{y^2} = \int_0^1 \frac{\partial e(\tau)}{\partial y} dx \Big/_{y=q} = 0.$$

Hieraus ist schließlich

$$\frac{3}{\pi} \iint_{\mathfrak{F}_1} e(\tau) \frac{dx\,dy}{y^2} = \delta(e) \tag{52}$$

zu entnehmen.

Wir wenden uns nun der Untersuchung von

$$R(s; e, g) = \frac{1}{\sqrt{\pi}} \iint_{\mathfrak{F}_1} \xi_2(s, \tau; g)\, \overline{e(\tau)} \frac{dx\,dy}{y^2} \tag{53}$$

zu. Hierin tragen wir die in $\mathfrak{F}_1$ gleichmäßig konvergente Entwicklung (39) für $\nu = 2$ ein und erhalten nach gliedweiser Integration

$$\begin{aligned} R(s; e, g) &= \frac{\Gamma(2s)}{\sqrt{\pi}\,(2\pi)^{2s}} \sum_{\{T\}>0} \frac{a(T)}{\varepsilon(T)} \sum_{|U|=1} \iint_{\mathfrak{F}_1} \{Sp(T[U]\,Y_1)\}^{-2s}\, \overline{e(\tau)} \frac{dx\,dy}{y^2} \\ &= \frac{2\,\Gamma(2s)}{\sqrt{\pi}\,(2\pi)^{2s}} \sum_{\{T\}>0} \frac{a(T)}{\varepsilon(T)} \int_0^\infty \int_{-\infty}^\infty \{Sp(T\,Y_1)\}^{-2s}\, \overline{e(\tau)} \frac{dx\,dy}{y^2}. \end{aligned} \tag{54}$$

$U$ durchläuft in der ersten Summe alle ganzzahligen Matrizen mit der Determinante 1. Es bleibt noch

$$\Omega(s, T) = \int_0^\infty \int_{-\infty}^\infty \{Sp(T\,Y_1)\}^{-2s}\, \overline{e(\tau)} \frac{dx\,dy}{y^2} \tag{55}$$

zu berechnen. Setzt man

$$\tau_1 = x_1 + i y_1 = |T|^{-\frac{1}{2}} (t_0 \tau + t_1),$$

so wird

$$Sp(T\,Y_1) = |T|^{\frac{1}{2}} \frac{x_1^2 + y_1^2 + 1}{y_1}, \quad \tau = \frac{|T|^{\frac{1}{2}} \tau_1 - t_1}{t_0}.$$

Die Variablentransformation $\tau \to \tau_1$ liefert also

$$\Omega(s, T) = |T|^{-s} \int_0^\infty \int_{-\infty}^\infty \left(\frac{y}{x^2 + y^2 + 1}\right)^{2s} \omega(\tau, T) \frac{dx\,dy}{y^2} \tag{56}$$

mit

$$\omega(\tau, T) = \overline{e\left(\frac{|T|^{\frac{1}{2}} \tau - t_1}{t_0}\right)}, \tag{57}$$

nachdem an Stelle von $\tau_1$ wieder $\tau$ geschrieben ist. Wir führen in (56) hyperbolische Polarkoordinaten $\varrho$, $\vartheta$ zum Mittelpunkt $\tau = i$ an Stelle von $x, y$ ein. Es gelten die Beziehungen:

$$\frac{\tau - i}{\tau + i} = \xi + i\eta \, (\xi, \eta \text{ reell}), \quad \xi = \varkappa \cos\vartheta, \quad \eta = \varkappa \sin\vartheta \qquad (\varkappa \geqq 0),$$

$$\varkappa = \mathfrak{Tg}\,\frac{\varrho}{2}, \quad \frac{x^2 + y^2 + 1}{y} = 2\,\mathfrak{Cof}\,\varrho, \quad \frac{dx\,dy}{y^2} = \mathfrak{Sin}\,\varrho\, d\varrho\, d\vartheta.$$

Die Wellengleichung (43), der $\omega = \omega(\tau, T)$ genügt, transformiert sich in

$$(58) \qquad \frac{1}{\mathfrak{Sin}\,\varrho}\left[\frac{\partial}{\partial\varrho}\left(\mathfrak{Sin}\,\varrho\,\frac{\partial\omega}{\partial\varrho}\right) + \frac{1}{\mathfrak{Sin}\,\varrho}\cdot\frac{\partial^2\omega}{\partial\vartheta^2}\right] + (\tfrac{1}{4} + r^2)\,\omega = 0$$

und (56) geht über in

$$(59) \qquad \Omega(s, T) = 4^{-s}\,|T|^{-s}\int_0^{2\pi}\int_0^{\infty}(\mathfrak{Cof}\,\varrho)^{-2s}\,\omega\,\mathfrak{Sin}\,\varrho\, d\varrho\, d\vartheta.$$

Wir entwickeln $\omega$ in eine Fourierreihe:

$$(60) \qquad \omega = \sum_{n=-\infty}^{\infty}\omega_n(\varrho)\, e^{2\pi i n\vartheta}.$$

Die Koeffizienten $\omega_n(\varrho)$ genügen gewissen Differentialgleichungen, die aus (58) zu gewinnen sind. $\omega_0(\varrho)$ insbesondere ist Lösung der Legendreschen Differentialgleichung

$$\frac{d}{dz}\left[(1 - z^2)\frac{d\omega_0}{dz}\right] - (\tfrac{1}{4} + r^2)\,\omega_0 = 0,$$

wenn $z = \mathfrak{Cof}\,\varrho$ gesetzt wird. Da $\omega_0$ für $\varrho = 0$ noch regulär ist, so ergibt sich nach[7]) Kap. IV., § 4, (6a)

$$\omega_0(\varrho) = c\,\mathfrak{P}_{-\frac{1}{2}+ir}(\mathfrak{Cof}\,\varrho).$$

Die Konstante $c$ hat den Wert $\omega_0(0)$. Damit wird

$$\omega(i, T) = \omega_{\varrho=0} = \lim_{\varrho\to 0}\frac{1}{2\pi}\int_0^{2\pi}\omega\, d\vartheta = \lim_{\varrho\to 0}\omega_0(\varrho) = c.$$

Wir tragen die Entwicklung (60) in (59) ein und führen die Integration über $\vartheta$ gliedweise aus. Offenbar liefert nur $\omega_0$ einen von 0 verschiedenen Beitrag zum Integral. Nach der Substitution $z = \mathfrak{Cof}\,\varrho$ erhält man

$$(61) \qquad \Omega(s, T) = 2\pi\,\omega(i, T)\,4^{-s}\,|T|^{-s}\int_1^{\infty} z^{-2s}\,\mathfrak{P}_{-\frac{1}{2}+ir}(z)\, dz.$$

Trägt man die Darstellung

$$K_{ir}(t) = \sqrt{\frac{\pi}{2}}\, t^{\frac{1}{2}}\int_1^{\infty} e^{-tz}\,\mathfrak{P}_{-\frac{1}{2}+ir}(z)\, dz \qquad \text{(s. }^{7}\text{) Kap. IV, § 5, Abschnitt f)}$$

in

$$2^{2s-\frac{5}{2}}\,\Gamma\left(s - \frac{ir}{2} - \frac{1}{4}\right)\Gamma\left(s + \frac{ir}{2} - \frac{1}{4}\right) = \int_0^{\infty} K_{ir}(t)\, t^{2s-\frac{3}{2}}\, dt \qquad \text{(s. }^{8}\text{) 13·21, (8))}$$

ein und vertauscht die Reihenfolge der Integrationen, so enthält man

---

[7]) Magnus, W., u. F. Oberhettinger: Formeln und Sätze für die speziellen Funktionen der mathematischen Physik. 2. Auflage. Berlin-Göttingen-Heidelberg: Springer-Verlag 1948.

[8]) Watson, G. N.: A treatise on the theory of Bessel functions. Cambridge 1922.

$$\int_1^\infty z^{-2s}\mathfrak{P}_{-\frac{1}{2}+ir}(z)\,dz = \frac{4^{s-1}}{\sqrt{\pi}}\,\frac{\Gamma\left(s-\frac{ir}{2}-\frac{1}{4}\right)\Gamma\left(s+\frac{ir}{2}-\frac{1}{4}\right)}{\Gamma(2s)}.$$

Die Berechnung von $\Omega$ ist damit durchgeführt:

$$(62)\qquad \Omega(s,T) = \frac{\sqrt{\pi}}{2}\,\omega(i,T)\,|T|^{-s}\,\frac{\Gamma\left(s-\frac{ir}{2}-\frac{1}{4}\right)\Gamma\left(s+\frac{ir}{2}-\frac{1}{4}\right)}{\Gamma(2s)}.$$

Es erweist sich als zweckmäßig, $\overline{e(\tau)}$ als Funktion von $Y$ darzustellen:

$$(63)\qquad e^*(Y) = \overline{e(\tau)}.$$

Alsdann wird

$$\omega(i,T) = \overline{e\left(\frac{|T|^{\frac{1}{2}}i - t_1}{t_0}\right)} = e^*(T^{-1}).$$

Die zu bestimmenden Funktionen (54) erscheinen nunmehr in ihrer endgültigen Gestalt:

$$(64)\qquad R(s;e,g) = (2\pi)^{-2s}\,\Gamma\left(s-\frac{ir}{2}-\frac{1}{4}\right)\Gamma\left(s+\frac{ir}{2}-\frac{1}{4}\right)\varphi(s;e,g).$$

Dabei sind

$$(65)\qquad \varphi(s;e,g) = \sum_{\{T\}>0}\frac{a(T)\,e^*(T^{-1})}{\varepsilon(T)\,|T|^s}$$

die der Modulform $g$ in natürlicher Weise zugeordneten DIRICHLETschen Reihen.

## § 2. Bestimmung der Funktionalgleichungen.

Eine Funktionalgleichung für $\varphi_1(s,\tau;g) + \varphi_2(s,\tau;g)$ läßt sich leicht bestimmen. Setzt man

$$(66)\quad \xi(s,\tau;g) = \xi_1(s,\tau;g) + \xi_2(s,\tau;g) = (2\pi)^{-2s}\,\Gamma(2s)\,[\varphi_1(s,\tau;g) + \varphi_2(s,\tau;g)],$$

so ergibt sich in der üblichen Weise (s. [2])) mit Hilfe von (30) und (31)

$$\begin{aligned}\xi(s,\tau;g) &= \int_0^\infty (f(uY_1) - a_0)\,u^{2s-1}\,du\\ (67)\qquad &= \int_1^\infty (f(uY_1) - a_0)\,u^{2s-1}\,du + (-1)^k\int_1^\infty (f(uY_1) - a_0)\,u^{2(k-s)-1}\,du\\ &\quad -\frac{a_0}{2s} - \frac{(-1)^k a_0}{2(k-s)},\end{aligned}$$

woraus

$$(68)\qquad \xi(k-s,\tau;g) = (-1)^k\,\xi(s,\tau;g)$$

erhellt. Aus (67) kann der analytische Charakter von $\varphi_2(s,\tau;g)$ bestimmt werden; denn $\varphi_1(s,\tau,g)$ setzt sich aus den Faktoren $\varphi^*(2s,g^*)$ und $G(\tau.\,2s)$ zusammen, von denen der erste durch die HECKEsche Theorie [2]) beherrscht wird, während für den zweiten in [9]) eine geeignete Reihenentwicklung angegeben worden ist. Der Vollständigkeit halber fügen wir noch die Formeln an:

[9]) MAASS, H.: Math. Ann. **121**, 141 (1949).

$$\xi^*(s, g^*) = (2\pi)^{-s}\,\Gamma(s)\,\varphi^*(s, g^*)$$

$$(69)\qquad = \int_1^\infty f_1^*(u)\,u^{s-1}\,du + i^k \int_1^\infty f_1^*(u)\,u^{k-s-1}\,du - \frac{a_0}{s} - \frac{i^k a_0}{k-s},$$

$$(70)\qquad \xi^*(k-s, g^*) = i^k\,\xi^*(s, g^*),$$

$$(71)\qquad \eta(s, \tau) = \pi^{-s}\,\Gamma(s)\,\zeta(2s)\,G(\tau, s),$$

$$(72)\qquad \eta(1-s, \tau) = \eta(s, \tau).$$

Diese Funktionalgleichung ist evident auf Grund der Entwicklung

$$\eta(s, \tau) = 2\pi^{-s}\,\Gamma(s)\,\zeta(2s)\,y^s + 2\pi^{\frac{1}{2}-s}\,\Gamma(s-\tfrac{1}{2})\,\zeta(2s-1)\,y^{1-s}$$

$$(73)\qquad + 2\sum_{n\neq 0}\left\{\sum_{d_1 d_2 = n} |d_1|^{\frac{1}{2}-s}\,|d_2|^{s-\frac{1}{2}}\right\} y^{\frac{1}{2}}\,K_{s-\frac{1}{2}}(2\pi|n|\,y)\,e^{2\pi i n x},$$

wenn hier noch die Funktionalgleichung der RIEMANNschen Zetafunktion $\zeta(s)$ berücksichtigt wird.

Unser eigentliches Ziel ist jedoch, Funktionalgleichungen für die Funktionen $\varphi(s; e, g)$ zu finden. Zu dem Zweck wird

$$R(s; e, g) = \frac{1}{\sqrt{\pi}} \iint_{\mathfrak{F}_1} \xi_2(s, \tau; g)\,\overline{e(\tau)}\,\frac{dx\,dy}{y^2} = \frac{1}{\sqrt{\pi}} \int_0^\infty \iint_{\mathfrak{F}_1} f_2(u\,Y_1)\,\overline{e(\tau)}\,u^{2s-1}\,du\,\frac{dx\,dy}{y^2}$$

umgeformt. Die Integration über $u$ unterbrechen wir an der Stelle $u = 1$. Das endliche Integral über $u$ von 0 bis 1 wird mit Hilfe der Substitution $u \to \frac{1}{u}$ in ein solches von 1 bis $\infty$ verwandelt. Beachtet man die Transformationsformeln (30), (31) und die Mittelwertgleichung (52), so ergibt sich für $R$ die Zerlegung

$$(74)\qquad R(s; e, g) = \frac{1}{\sqrt{\pi}} \int_1^\infty \iint_{\mathfrak{F}_1} f_2(Y)\,\overline{e(\tau)}\,u^{2s-1}\,du\,\frac{dx\,dy}{y^2}$$

$$+ \frac{(-1)^k}{\sqrt{\pi}} \int_1^\infty \iint_{\mathfrak{F}_1} f_2(Y)\,\overline{e(\tau)}\,u^{2(k-s)-1}\,du\,\frac{dx\,dy}{y^2} - \left(\frac{1}{s} + \frac{(-1)^k}{k-s}\right)\frac{\sqrt{\pi}\,a_0\,\delta(e)}{6}$$

$$+ \frac{1}{\sqrt{\pi}} \int_1^\infty \iint_{\mathfrak{F}_1} \left[(-1)^k\,|Y|^k\,f_1(Y) - f_1(Y^{-1})\right]\overline{e(\tau)}\,u^{-2s-1}\,du\,\frac{dx\,dy}{y^2}.$$

Um hier eine Invarianz gegenüber der Substitution $s \to k - s$ zu erkennen, müssen wir das letzte Integral geeignet umformen. Zugleich ist damit die Frage der analytischen Fortsetzung von $\varphi(s; e, g)$ zu klären. Ist doch die vorgenommene Umformung von $R$ nur für $s$-Werte mit hinreichend großem Realteil erlaubt! Wir können das letzte Integral offenbar als Grenzwert des Ausdrucks

$$(75)\qquad J(p, q) = \frac{1}{\sqrt{\pi}} \int_1^p \iint_{\mathfrak{B}_q} \left[(-1)^k\,u^{2k}\,f_1(Y) - f_1(Y^{-1})\right]\overline{e(\tau)}\,u^{-2s-1}\,du\,\frac{dx\,dy}{y^2}$$

für $p, q \to \infty$ darstellen. Dabei bezeichnet wie bisher $\mathfrak{B}_q$ den Bereich $0 < y \leq q$, $|2x| \leq 1$, $x^2 + y^2 \geq 1$. Sei

$$
(76)\quad \begin{aligned}
J_1(p,q) &= \frac{(-1)^k}{\sqrt{\pi}} \int_1^p \iint_{\mathfrak{B}_q} f_1(Y)\,\overline{e(\tau)}\,u^{2(k-s)-1}\,du\,\frac{dx\,dy}{y^2},\\
J_2(p,q) &= \frac{1}{\sqrt{\pi}} \int_1^p \iint_{\mathfrak{B}_q} f_1(Y^{-1})\,\overline{e(\tau)}\,u^{-2s-1}\,du\,\frac{dx\,dy}{y^2}.
\end{aligned}
$$

Dann wird

$$
(77)\qquad J(p,q) = J_1(p,q) - J_2(p,q).
$$

Die Integrale (76) werden nun mit Hilfe der Entwicklung (29) berechnet. Zunächst bestimmen wir zu jedem Paar teilerfremder ganzer Zahlen $c$, $d$ eine Modulsubstitution $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, die den Fundamentalbereich $\mathfrak{F}_1$ in den Streifen $|2x| \leqq 1$ abbildet. $U$ ist durch $c$, $d$ eindeutig bestimmt. Wir setzen

$$
(78)\qquad \mathfrak{S}_q = \sum_U U(\mathfrak{B}_q),
$$

wobei $U$ ein volles System von Modulsubstitutionen der angegebenen Art durchlaufen möge, so daß die zweiten Zeilen zweier Substitutionen $U$ sich nicht nur um einen Faktor $\pm 1$ unterscheiden. Damit ergibt sich

$$
(79)\quad \begin{aligned}
J_1(p,q) &= \frac{(-1)^k}{2\sqrt{\pi}} \sum_{(c,d)=1} \int_1^p \iint_{\mathfrak{B}_q} f_1^*\left(u\,\frac{|c\tau+d|^2}{y}\right)\overline{e(\tau)}\,u^{2(k-s)-1}\,du\,\frac{dx\,dy}{y^2}\\
&= \frac{(-1)^k}{\sqrt{\pi}} \int_1^p \iint_{\mathfrak{S}_q} f_1^*\left(\frac{u}{y}\right)\overline{e(\tau)}\,u^{2(k-s)-1}\,du\,\frac{dx\,dy}{y^2}.
\end{aligned}
$$

Sind $\tau$ und $\tau_1$ Punkte des Streifens $|2x| \leqq 1$, $y > 0$, die bezüglich der Modulgruppe äquivalent sind, so stimmen $y = \mathfrak{Im}\,\tau$ und $y_1 = \mathfrak{Im}\,\tau_1$ überein oder es ist $y\,y_1 \leqq 1$. Sei nun $\frac{1}{q} \leqq y \leqq q$, $\tau_1 \subset \mathfrak{F}_1$. Dann ist entweder $y_1 = y \leqq q$ oder $y_1 \leqq \frac{1}{y} \leqq q$, jedenfalls also $\tau_1 \subset \mathfrak{B}_q$, mithin $\tau \subset \mathfrak{S}_q$. Der Bereich

$$
\frac{1}{q} \leqq y \leqq q, \quad |2x| \leqq 1
$$

ist also in $\mathfrak{S}_q$ enthalten. Es sei $M$ eine Schranke für $e(\tau)$:

$$
|e(\tau)| \leqq M \quad \text{für} \quad y > 0.
$$

Wir entwickeln die Eigenfunktion in eine Fourierreihe. Nach (51) kann

$$
e(\tau) = \delta(e) + \sum_{n \neq 0} c_n(y)\,e^{2\pi i n x}
$$

gesetzt werden. Sodann ergibt sich mit einer gewissen Funktion $C_q(v)$, die der Abschätzung $|C_q(v)| \leqq M$ genügt,

$$\begin{aligned}
(80)\qquad & \iint\limits_{\mathfrak{S}_q} e^{-2\pi t\frac{u}{y}}\,\overline{e(\tau)}\,\frac{dx\,dy}{y^2} \\
&= \iint\limits_{\substack{\frac{1}{q}\le y\le q\\ |2x|\le 1}} e^{-2\pi t\frac{u}{y}}\,\overline{e(\tau)}\,\frac{dx\,dy}{y^2} + \iint\limits_{\substack{\mathfrak{S}_q\\ qy\le 1}} e^{-2\pi t\frac{u}{y}}\,\overline{e(\tau)}\,\frac{dx\,dy}{y^2} \\
&= \delta(e)\int\limits_{\frac{1}{q}}^{q} e^{-2\pi t\frac{u}{y}}\,\frac{dy}{y^2} + \iint\limits_{\substack{\mathfrak{S}_q\\ qy\le 1}} e^{-2\pi t\frac{u}{y}}\,\overline{e(\tau)}\,\frac{dx\,dy}{y^2} \\
&= \delta(e)\,\frac{e^{-2\pi t\frac{u}{y}}}{2\pi t u}\Bigg|_{y=\frac{1}{q}}^{y=q} + C_q(tu)\,\frac{e^{-2\pi t\frac{u}{y}}}{2\pi t u}\Bigg|_{y\to 0}^{y=\frac{1}{q}} \\
&= \frac{1}{2\pi t u}\left\{\delta(e)\,e^{-2\pi t\frac{u}{q}} - C_q^*(tu)\,e^{-2\pi t u q}\right\}.
\end{aligned}$$

Dabei ist $C_q^*(v) = \delta(e) - C_q(v)$ gesetzt worden. Man erkennt sofort, daß $|C_q^*(v)| \le M$ ist. Wir wenden die abgeleitete Formel an, um $J_1(p, q)$ zu berechnen:

$$\begin{aligned}
J_1(p,q) &= \frac{(-1)^k}{\sqrt{\pi}}\sum_{t=1}^{\infty} a_t \int\limits_1^p \iint\limits_{\mathfrak{S}_q} e^{-2\pi t\frac{u}{y}}\,\overline{e(\tau)}\,\frac{dx\,dy}{y^2}\,u^{2(k-s)-1}\,du \\
(81)\qquad &= \frac{(-1)^k}{\sqrt{\pi}}\sum_{t=1}^{\infty}\frac{a_t}{2\pi t}\int\limits_1^p \left\{\delta(e)\,e^{-2\pi t\frac{u}{q}} - C_q^*(tu)\,e^{-2\pi t u q}\right\} u^{2(k-s-1)}\,du.
\end{aligned}$$

Analog verläuft die Rechnung für das zweite der Integrale (76):

$$\begin{aligned}
J_2(p,q) &= \frac{1}{2\sqrt{\pi}}\sum_{(c,d)=1}\int\limits_{\frac{1}{p}}^{1}\iint\limits_{\mathfrak{B}_q} f_1^*\!\left(u\,\frac{|c\tau+d|^2}{y}\right)\overline{e(\tau)}\,u^{2s-1}\,du\,\frac{dx\,dy}{y^2} \\
(82)\qquad &= \frac{1}{\sqrt{\pi}}\int\limits_{\frac{1}{p}}^{1}\iint\limits_{\mathfrak{S}_q} f_1^*\!\left(\frac{u}{y}\right)\overline{e(\tau)}\,u^{2s-1}\,du\,\frac{dx\,dy}{y^2} \\
&= \frac{1}{\sqrt{\pi}}\sum_{t=1}^{\infty} a_t\int\limits_{\frac{1}{p}}^{1}\iint\limits_{\mathfrak{S}_q} e^{-2\pi t\frac{u}{y}}\,\overline{e(\tau)}\,\frac{dx\,dy}{y^2}\,u^{2s-1}\,du \\
&= \frac{1}{\sqrt{\pi}}\sum_{t=1}^{\infty}\frac{a_t}{2\pi t}\int\limits_{\frac{1}{p}}^{1}\left\{\delta(e)\,e^{-2\pi t\frac{u}{q}} - C_q^*(tu)\,e^{-2\pi t u q}\right\} u^{2(s-1)}\,du.
\end{aligned}$$

Wir setzen zur Abkürzung

$$(83)\qquad h(y) = \sum_{t=1}^{\infty}\frac{a_t}{2\pi t}\,e^{-2\pi t y}.$$

Dann wird

(84) $$h'(y) = -f_1^*(y)$$

und

(85) $$J(p,q) = \frac{(-1)^k \delta(e)}{\sqrt{\pi}} \int_1^p h\left(\frac{u}{q}\right) u^{2(k-s-1)}\, du - \frac{\delta(e)}{\sqrt{\pi}} \int_{\frac{1}{p}}^1 h\left(\frac{u}{q}\right) u^{2(s-1)}\, du$$
$$- \frac{(-1)^k}{\sqrt{\pi}} \sum_{t=1}^{\infty} \frac{a_t}{2\pi t} \int_1^p C_q^*(tu)\, e^{-2\pi t u q} u^{2(k-s-1)}\, du$$
$$+ \frac{1}{\sqrt{\pi}} \sum_{t=1}^{\infty} \frac{a_t}{2\pi t} \int_{\frac{1}{p}}^1 C_q^*(tu)\, e^{-2\pi t u q} u^{2(s-1)}\, du.$$

Der Beitrag der letzten beiden Summen zum Integral $J(p, q)$ wird unter der Voraussetzung $\sigma = \Re s > k$ abgeschätzt durch

$$\frac{M}{2\pi\sqrt{\pi}} \sum_{t=1}^{\infty} \frac{|a_t|}{t} \left\{ \int_1^{\infty} e^{-2\pi t u q}\, du + \int_0^1 e^{-2\pi t u q} u^{2(\sigma-1)}\, du \right\}$$
$$\leq \frac{M}{2\pi\sqrt{\pi}} \sum_{t=1}^{\infty} \frac{|a_t|}{t} \int_0^{\infty} e^{-2\pi t u q} u^{2(\sigma-1)}\, du = \frac{M\,\Gamma(2\sigma-1)}{(2\pi)^{2\sigma}\sqrt{\pi}\, q^{2\sigma-1}} \sum_{t=1}^{\infty} \frac{|a_t|}{t^{2\sigma}}$$
$$= O\left(\frac{1}{q}\right) \quad \text{für } q \to \infty \text{ gleichmäßig in } p.$$

Demnach ist

(86) $$\lim_{p,q\to\infty} J(p,q) = \lim_{p,q\to\infty} J^*(p,q),$$

wenn wir

(87) $$J^*(p,q) = \frac{(-1)^k \delta(e)}{\sqrt{\pi}} \int_1^p h\left(\frac{u}{q}\right) u^{2(k-s-1)}\, du - \frac{\delta(e)}{\sqrt{\pi}} \int_{\frac{1}{p}}^1 h\left(\frac{u}{q}\right) u^{2(s-1)}\, du$$

setzen. Partielle Integration führt auf

(88) $$J^*(p,q) = \frac{(-1)^k \delta(e)}{\sqrt{\pi}} \left. \frac{h\left(\frac{u}{q}\right) u^{2(k-s)-1}}{2(k-s)-1} \right|_{u=1}^{u=p} - \frac{\delta(e)}{\sqrt{\pi}} \left. \frac{h\left(\frac{u}{q}\right) u^{2s-1}}{2s-1} \right|_{u=\frac{1}{p}}^{u=1}$$
$$+ \frac{(-1)^k \delta(e)}{\sqrt{\pi}\, q(2k-2s-1)} \int_1^p f_1^*\left(\frac{u}{q}\right) u^{2(k-s)-1}\, du - \frac{\delta(e)}{\sqrt{\pi}\, q(2s-1)} \int_{\frac{1}{p}}^1 f_1^*\left(\frac{u}{q}\right) u^{2s-1}\, du.$$

Wir vollziehen den Grenzübergang $p \to \infty$ und erhalten

(89) $$J^*(q) = \lim_{p\to\infty} J^*(p,q) = -\frac{(-1)^k \delta(e)\, h\left(\frac{1}{q}\right)}{\sqrt{\pi}\,(2k-2s-1)} - \frac{\delta(e)\, h\left(\frac{1}{q}\right)}{\sqrt{\pi}\,(2s-1)}$$
$$+ \frac{(-1)^k \delta(e)}{\sqrt{\pi}\, q(2k-2s-1)} \int_1^{\infty} f_1^*\left(\frac{u}{q}\right) u^{2(k-s)-1}\, du - \frac{\delta(e)}{\sqrt{\pi}\, q(2s-1)} \int_0^1 f_1^*\left(\frac{u}{q}\right) u^{2s-1}\, du.$$

Nach (32) ist

$$f_1^*\left(\frac{u}{q}\right) = i^k \left(\frac{q}{u}\right)^k f_1^*\left(\frac{q}{u}\right) + \left(i^k \left(\frac{q}{u}\right)^k - 1\right) a_0.$$

Damit wird

$$J^*(q) = -\frac{\delta(e)}{\sqrt{\pi}}\, h\left(\frac{1}{q}\right)\left(\frac{1}{2s-1} + \frac{(-1)^k}{2(k-s)-1}\right)$$

$$+ \frac{(-i)^k \delta(e)\, q^{k-1}}{\sqrt{\pi}\,(2k-2s-1)} \int_1^\infty f_1^*\left(\frac{q}{u}\right) u^{k-2s-1}\, du - \frac{i^k \delta(e)\, q^{k-1}}{\sqrt{\pi}\,(2s-1)} \int_0^1 f_1^*\left(\frac{q}{u}\right) u^{-k+2s-1}\, du$$

$$- \frac{(-i)^k \delta(e)\, q^{k-1} a_0}{\sqrt{\pi}\,(2k-2s-1)(k-2s)} + \frac{(-1)^k \delta(e)\, a_0}{2q\sqrt{\pi}\,(2k-2s-1)(k-s)}$$

$$- \frac{i^k \delta(e)\, q^{k-1} a_0}{\sqrt{\pi}\,(2s-1)(2s-k)} + \frac{\delta(e)\, a_0}{2q\sqrt{\pi}\,(2s-1)s}$$

$$= -\frac{\delta(e)}{\sqrt{\pi}}\, h\left(\frac{1}{q}\right)\left(\frac{1}{2s-1} + \frac{(-1)^k}{2(k-s)-1}\right)$$

$$+ \frac{(-i)^k \delta(e)\, q^{2(k-s)-1}}{\sqrt{\pi}\,(2k-2s-1)} \int_0^q f_1^*(u)\, u^{2s-k-1}\, du - \frac{i^k \delta(e)\, q^{2s-1}}{\sqrt{\pi}\,(2s-1)} \int_q^\infty f_1^*(u)\, u^{k-2s-1}\, du$$

$$+ \frac{i^k \delta(e) q^{k-1} a_0}{\sqrt{\pi}\,(2s-k)}\left(\frac{(-1)^k}{2(k-s)-1} - \frac{1}{2s-1}\right) + \frac{\delta(e)\, a_0}{2q\sqrt{\pi}}\left(\frac{(-1)^k}{(2k-2s-1)(k-s)} + \frac{1}{(2s-1)s}\right).$$

Wir ziehen die für $q \to \infty$ ersichtlich gegen 0 gehenden Glieder heraus, bilden also

$$J(q) = -\frac{\delta(e)}{\sqrt{\pi}}\, h\left(\frac{1}{q}\right)\left(\frac{1}{2s-1} + \frac{(-1)^k}{2(k-s)-1}\right) \tag{90}$$
$$+ \frac{i^k \delta(e)\, q^{k-1} a_0}{\sqrt{\pi}\,(2s-k)}\left(\frac{(-1)^k}{2(k-s)-1} - \frac{1}{2s-1}\right).$$

Es ist dann

$$\lim_{p,q\to\infty} J(p,q) = \lim_{q\to\infty} J^*(q) = \lim_{q\to\infty} J(q). \tag{91}$$

Berücksichtigen wir die Identität

$$\frac{1}{2s-k}\left(\frac{(-1)^k}{2(k-s)-1} - \frac{1}{2s-1}\right) = \frac{1}{k-1}\left(\frac{(-1)^k}{2(k-s)-1} + \frac{1}{2s-1}\right) + \frac{(-1)^k-1}{k-1}\,\frac{1}{2s-k}$$

für $k > 1$,

so folgt wegen $a_0((-1)^k - 1) = 0$, daß

$$J(q) = \frac{\delta(e)}{\sqrt{\pi}}\left(\frac{1}{2s-1} + \frac{(-1)^k}{2(k-s)-1}\right)\left(-h\left(\frac{1}{q}\right) + \frac{i^k a_0 q^{k-1}}{k-1}\right) \tag{92}$$

ist. Es bleibt der Grenzwert der letzten Klammer zu bestimmen. Ausgehend von der Formel

$$e^{-2\pi t y} = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \Gamma(s)\,(2\pi t y)^{-s}\, ds$$

finden wir

$$h(y) = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \sum_{t=1}^{\infty} \frac{a_t\, \Gamma(s)}{2\pi t\,(2\pi t y)^s}\, ds = \frac{1}{4\pi^2 i} \int_{\sigma-i\infty}^{\sigma+i\infty} \frac{\varphi^*(s+1, g^*)\, \Gamma(s)}{(2\pi y)^s}\, ds,$$

wenn wir $\sigma$ hinreichend groß wählen. Wir verschieben die Integrationsgerade parallel bis zum Punkt $s = -\frac{1}{2}$. Da $\Gamma(s)$ in $s = 0$ und $\varphi^*(s+1, g^*)$

in $s = k - 1$ je einen Pol erster Ordnung haben[2]), so wird nach dem Residuensatz

$$(93)\quad h(y) = \frac{1}{2\pi}\left\{\frac{1}{2\pi i}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty}\frac{\varphi^*(s+1,g^*)\,\Gamma(s)}{(2\pi y)^s}\,ds + \varphi^*(1,g^*) + \frac{\alpha\,\Gamma(k-1)}{(2\pi y)^{k-1}}\right\}.$$

Dabei bezeichnet $\alpha$ das Residuum von $\varphi^*(s, g^*)$ in $s = k$. Man stellt leicht fest, daß das letzte Integral mit $y$ gegen 0 konvergiert. Infolgedessen ist

$$(94)\qquad -\frac{1}{2\pi}\,\varphi^*(1,g^*) = \lim_{q\to\infty}\left(-h\left(\frac{1}{q}\right) + \frac{\alpha\,\Gamma(k-1)\,q^{k-1}}{(2\pi)^k}\right).$$

Aus der Heckeschen Theorie[2]) entnehmen wir noch die Beziehung

$$(95)\qquad a_0 = (-1)^k a_0 = i^k (2\pi)^{-k}\,\Gamma(k)\,\alpha.$$

Alsdann wird

$$\lim_{q\to\infty}\left(-h\left(\frac{1}{q}\right) + \frac{i^k a_0 q^{k-1}}{k-1}\right) = -\frac{1}{2\pi}\,\varphi^*(1,g^*)$$

und

$$(96)\qquad \lim_{q\to\infty} J(q) = -\frac{\delta(e)\,\varphi^*(1,g^*)}{2\pi\sqrt{\pi}}\left(\frac{1}{2s-1} + \frac{(-1)^k}{2(k-s)-1}\right).$$

Diese Beziehung gilt nicht nur für $k > 1$, sondern auch für $k = 1$, weil dann beide Seiten der Gleichung verschwinden. Aus (74) ergibt sich damit die gewünschte Darstellung

$$(97)\qquad \begin{aligned} R(s; e, g) = {} & \frac{1}{\sqrt{\pi}}\int_1^\infty \iint_{\mathfrak{F}_1} f_2(Y)\,\overline{e(\tau)}\,u^{2s-1}\,du\,\frac{dx\,dy}{y^2} \\ & + \frac{(-1)^k}{\sqrt{\pi}}\int_1^\infty \iint_{\mathfrak{F}_1} f_2(Y)\,\overline{e(\tau)}\,u^{2(k-s)-1}\,du\,\frac{dx\,dy}{y^2} \\ & - \frac{\sqrt{\pi}\,a_0\,\delta(e)}{6}\left(\frac{1}{s} + \frac{(-1)^k}{k-s}\right) - \frac{\delta(e)\,\varphi^*(1,g^*)}{2\pi\sqrt{\pi}}\left(\frac{1}{2s-1} + \frac{(-1)^k}{2(k-s)-1}\right). \end{aligned}$$

Sie ist gültig für alle $s$ und liefert die analytische Fortsetzung von $\varphi(s; e, g)$ in die ganze $s$-Ebene. $\varphi(s; e, g)$ ist abgesehen von endlich vielen Polen erster Ordnung überall regulär und genügt offenbar der Funktionalgleichung

$$(98)\qquad R(k-s; e, g) = (-1)^k R(s; e, g).$$

Anwenden lassen sich die vorstehenden Untersuchungen auf die Eisensteinreihen. Die Fourierkoeffizienten dieser Reihen haben, wie Siegel[1]) gezeigt hat, multiplikative Eigenschaften. Es wäre von Interesse, festzustellen, ob es sich hier um eine Besonderheit der Eisensteinreihen handelt oder ob auch anderen Modulformen zweiten Grades diese Eigenschaft zukommt. Diese Frage ist wegen der Beziehungen der Modulformen zu den Dirichletreihen und wegen möglicher zahlentheoretischer Anwendungen von erhöhter Bedeutung.

*(Eingegangen am 21. September 1949.)*

122. Band. **Inhalt:** 1. Heft.



Unveränderter Nachdruck 1970
Springer-Verlag Berlin Heidelberg New York
Gesamtherstellung: fotokop wilhelm weihert, Darmstadt